PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/85, 5/06, G01N 33/567, C07H 21/00 — A1

(11) International Publication Number: WO 98/42854

(43) International Publication Date: 1 October 1998 (01.10.98)

(21) International Application Number: PCT/US98/06093

(22) International Filing Date: 27 March 1998 (27.03.98)

(30) Priority Data:

60/042,543	27 March 1997 (27.03.97)	US
Not furnished	26 March 1998 (26.03.98)	US

(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(71) Applicant: THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY [US/US]; Suite 350, 900 Welch Road, Palo Alto, CA 94304–1850 (US).

(72) Inventors: BLAU, Helen, M.; 580 Cotton Street, Menlo Park, CA 94025 (US). SPICHER, Albert; 105 Amherst Avenue, Menlo Park, CA 94025 (US). GUICHERIT, Oivin; 28 Hamilton Road, Belmont, MA 02178 (US).

(74) Agents: BRENNAN, Sean, M. et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304–1018 (US).

(54) Title: FUNCTIONAL GENOMIC SCREEN FOR RNA REGULATORY SEQUENCES AND INTERACTING MOLECULES

(57) Abstract

Methods and compositions for the identification, characterization and isolation of regulatory RNA sequences are provided. Regulatory RNA sequences mediate post–transcriptional regulation in response to various environmental conditions, can be used to alter the level of expression of endogenous genes, and can be used to identify factors which interact with regulatory RNA sequences. The invention additionally provides improved vector systems for rapid screening, analysis, and tightly–regulated expression of regulatory RNA sequences.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	LS	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	LT	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	LU	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	LV	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	MC	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	MD	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	MG	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	MK	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	ML	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	MN	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	MR	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	MW	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	MX	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	NE	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	NL	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	NO	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	**KR**	Republic of Korea	PT	Portugal		
CU	Cuba	**KZ**	Kazakstan	RO	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	RU	Russian Federation		
DE	Germany	**LI**	Liechtenstein	SD	Sudan		
DK	Denmark	**LK**	Sri Lanka	SE	Sweden		
EE	Estonia	**LR**	Liberia	SG	Singapore		

FUNCTIONAL GENOMIC SCREEN FOR RNA REGULATORY SEQUENCES AND INTERACTING MOLECULES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Patent Application Serial No. 60/042,543 filed March 27, 1997 and to U.S. Patent Application Attorney Docket Number 28600-20209.00, filed March 26, 1998.

STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH

Some of the research described herein was conducted under grant No. AG 09521 from the National Institute of Aging. The United States Government may have certain rights in this invention.

TECHNICAL FIELD

This invention is in the field of molecular biology. More particularly, it is in the field of RNA regulatory sequences affecting post-transcriptional processes, retroviral vectors and regulated expression systems. It is also in the field of functional genomics, involving the screening and identification of novel RNA regulatory sequences and molecules with which they interact.

BACKGROUND

A complex regulatory circuitry underlies mammalian growth and differentiation. Although many tissue-specific regulators are transcription factors that bind DNA, others act indirectly. For example, regulators can sequester other regulators in intracellular compartments, bind cooperatively with ubiquitous transcription factors, modify these factors by changing their phosphorylation state, or facilitate their dimerization. Thus, an elucidation of growth and differentiation control pathways requires a combination of assays based on binding to known molecules with assays based on function.

The identification of regulatory elements required for correct expression of genes is an essential step in understanding their function. Most studies of the regulation of gene expression have focused on the initiation of transcription. However, it has been recently

recognized that, for many genes, post-transcriptional processes play a major role in determining when, where and at what level genes are expressed.

Transcriptional controls influence gene expression by determining rates of mRNA production, but post-transcriptional controls are equally important in that they determine the amount of protein produced from an mRNA. In mammals, post-transcriptional control appears to be especially important for a cell's response to changes in the environment, such as heat shock (Sierra *et al.* (1994), *Mol. Biol. Rep.* **19**:211-220), the availability of iron (Hentze *et al.* (1996) *Proc. Natl. Acad. Sci. USA* **93**:8175-8182), oxygen (Levy *et al.* (1996) *J. Biol Chem* **271**:2746-2753; and McGary *et al.* (1997) *J. Biol Chem* **272**:8628-8634), or growth factors (Amara *et al.* (1993) *Nucleic Acids Res* **21**:4803-4809). Post-transcriptional mechanisms may also serve to check and balance transcriptional regulation of gene expression. Although splicing (Nilsen (1998) in RNA Structure and Function, R. W. Simons and M. Grunberg-Manago, eds., Plainview, New-York: Cold Spring Harbor Laboratory Press, pp. 279-308), integration of selenocysteine (Huettenhofer *et al.* (1998) in "RNA Structure and Function," R. W. Simons and M. Grunberg-Manago, eds., Plainview, New-York: Cold Spring Harbor Laboratory Press, pp. 603-668), editing (Connell *et al.* (1998) in "RNA Structure and Function," R. W. Simons and M. Grunberg-Manago, eds., Plainview, New-York: Cold Spring Harbor Laboratory Press, pp. 641-667) and frameshifting (Atkins *et al.* (1996) in "Translational Control," J. W. B. Hershey, M. B. Mathews and N. Sonenberg, eds., Plainview, New-York: Cold Spring Harbor Laboratory Press, pp. 653-684) are documented post-transcriptional control mechanisms, two particularly well established mechanisms for modulating gene expression post-transcriptionally are alterations in mRNA stability and mRNA translation efficiency. Ross (1995) *Microbiol. Rev.* **59**:423-450; Izquierdo *et al.* (1997) *Mol. Cell. Biol.* **17**:5255-5268; and Yang (1997) *J. Biol. Chem.* **272**:15466-15473.

It is now clear that protein noncoding sequences, such as 3'-untranslated regions of messenger RNAs (3' UTRs), contain numerous regulatory elements that affect the transport, subcellular localization, translation and degradation of messenger RNAs (Jackson (1993) *Cell* **74**:9-14). Some 3' UTRs (riboregulators) may even act in trans to affect the expression of other genes (Rastinejad *et al.* (1993) *Cell* **72**:903-917; L'Ecuyer *et al.* (1995) *Proc. Natl. Acad. Sci USA* **92**:7520-7524; and Davis *et al.* (1996) *Proc. Natl. Acad. Sci. USA* **93**:508-513). As a result, riboregulators can have profound effects on cell proliferation, differentiation and cancer. Rastinejad *et al.* (1993) *Cell* **75**:1107-1117;

L'Ecuyer *et al, supra*; Jupe *et al.* (1996) *Cell Growth Differ.* 7:871-878; Chen *et al.* (1996) *Int. J. Rad. Biol.* **69**:385-395); Jupe *et al.* (1996) *Exp. Cell Res.* **224**:128-135; and Jan *et al.* (1997) *EMBO J.* **16**:6301-6313.

Regulation of actin genes by 3'UTRs has been reported. Expression of a stably transfected chimeric gene of the 5' regulatory region of β-actin linked to the 3'UTR of α-skeletal actin does not decline during differentiation, but increases like the endogenous α-skeletal actin gene (Sharp *et al.* (1989) *Gene* 80, 293-304). Conversely, when the 3'UTR of β-actin is linked to heterologous promoters, expression from these promoters, like the endogenous β-actin gene declines during differentiation (DePonti-Zilli *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85, 1389-1393). Blau *et al.* (1993) *J. Cell Biology Supplement* 17A: 58, reported effects on cell growth exerted by the 3' untranslated regions of muscle structural genes. See also, Rastinejad *et al.* (1993) *Cell* 72, 903-917; L'Ecuyer *et al., supra*; and Davis *et al.* (1996), *supra*.

In general, however, there is a paucity of well-characterized examples of gene regulation in vertebrates by UTRs, due to the cumbersome nature of the available tools. The major approach to the study of UTRs to date has been to introduce a UTR-reporter gene construct, by transient or stable transfection, into an established cell line. However, transient transfections are difficult to control and suffer from low efficiency, variable expression levels and limited persistence of expression; while production of stably transfected clones is both time-consuming and labor-intensive. Furthermore, stable transfections yield only a few clones after weeks to months of selection, and the clones obtained from stable transfections may not be representative, especially if the 3'UTR has an adverse effect on growth. Thus, a method that overcomes these problems would be of great value for functional genomic screening.

Highly conserved regions (HCRs) have been shown to exist within 3'UTRs, and it has been speculated that these HCRs may play a role in mRNA metabolism. Duret *et al.* (1993) *Nucleic Acids Res.* **21**:2315-2322. However, a rigorous test of the hypothesis that HCRs serve a regulatory function, as well as elucidation of the mechanism of action of a HCR, have been difficult to achieve, for several reasons. First, plasmid-based expression systems presently available suffer from low efficiency, variable expression level and limited persistence of expression. Second, since 3'UTRs invariably contain signals for mRNA 3' end formation, insertion of a 3' UTR into an expression vector will often cause premature transcription termination of transcripts important for the replication of the

vector, thus making it difficult to propagate vectors containing 3' UTRs. Consequently, the nature of most 3' UTR post-transcriptional regulatory elements and the processes in which they are involved remain poorly understood, especially in mammals. Thus, there is a need in the art for methods to identify 3'UTR sequences and elucidate their mechanisms of action in mammalian cells. Such methods require the ability to introduce and express wild-type and mutant 3'UTR sequences efficiently and systematically in mammalian cells.

Hofmann *et al.* (1996) *Proc. Natl. Acad. Sci. USA* **93**:5185-5190, describe rapid retroviral delivery of tetracycline-inducible genes in a single autoregulatory cassette. The structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance has been described. Hinrichs *et al.* (1994) *Science* **264**:418-420. The use of fluorescence histochemistry to monitor gene expression is presented in Mohler *et al.* (1996) *Proc. Natl. Acad. Sci. USA.* **93**:12423-12427.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide methods and compositions for the identification of RNA regulatory sequences. Such sequences can regulate the stability, localization, transport and/or translation of a mRNA molecule, or can regulate any other aspect of mRNA metabolism that will affect the levels or localization of the mRNA or its polypeptide product.

It is a further object to provide a method for high-throughput screening for RNA regulatory sequences, enabling screening of, for example, oligo dT-primed cDNA libraries and other sequences which contain signals for mRNA 3'-end formation.

It is yet a further object to provide methods and compositions for post-transcriptional regulation of the expression of a therapeutic gene. Post-transcriptional regulation encompasses both positive and negative responses to environmental conditions including, but not limited to, stress, mitogen concentration, temperature, small molecules, oxygen tension, *etc.* A positive response will result in an increase in production of a polypeptide encoded by a therapeutic gene; a negative response will result in a decrease in production of a polypeptide encoded by a therapeutic gene.

A further object of the invention is to provide compositions and methods for modulating the expression of genes related to pathogenesis Such modulation can include inhibition or activation of gene expression.. Such genes include, but are not limited to, oncogenes, viral genes, genes involved in apoptosis, genes involved in senescence, genes

involved in metastasis, genes involved in angiogenesis and genes involved in drug resistance.

It is another object of the invention to provide methods for identifying molecules which interact with RNA regulatory sequences. Such molecules include, but are not limited to, proteins, RNA molecules, DNA molecules, small molecules, drugs, and ions.

In accord with the aforementioned objects, the present invention provides compositions and methods for the identification, analysis and isolation and use of regulatory RNA sequences. These include, for example, post-transcriptional regulatory elements located in the untranslated regions (UTRs) of mRNAs, which can, in some cases, be identified by sequence conservation. RNA regulatory sequences also include sequences located in introns and 5' untranslated regions, protein-coding regions, and untranscribed, intergenic regions, or can comprise synthetic, non-naturally-occurring sequences, such as might be produced, for example, in a combinatorial library of random RNA sequences. Preferred are highly conserved regions of sequence (HCRs) located in 3'-untranslated regions (UTRs). Regulatory RNA sequences identified through the practice of the invention regulate various aspects of messenger RNA metabolism, including but not limited to translation, mRNA stability, mRNA transport, mRNA localization, mRNA processing, insertion of selenocysteine and ribosomal frameshifting during translation of mRNA. Regulation can be positive or negative, and can occur in response to any type of environmental condition including but not limited to mitogen stimulation, growth factor stimulation, tissue damage, ischemia, infection (viral or bacterial), variations in oxygen tension (*e.g.*, hypoxia), presence of cell-cycle regulators, stresses such as temperature (*e.g.*, heat shock and cold shock) and pressure; variations in ionic strength (*e.g.*, hyper-and hypotonicity, presence of specific ions, such as Ca^{2+}) and pH, presence of metal ions, differentiation factors, angiogenic factors, senescence factors, toxins, carcinogens, teratogens, metabolites, extracellular matrix components, drugs, small molecules, and hormones.

The invention further provides a method for high-throughput genomic screening to identify and characterize regulatory RNA sequences, utilizing a retroviral-based vector system in which post-transcriptional effects can be isolated from transcriptional effects and in which different post-transcriptional regulatory elements can be compared with one another in the absence of differential transcriptional effects.

The method, in one embodiment, utilizes two or more retrovirus-based vectors. A reporter vector comprises a reporter gene encoding a detectable marker and optionally a transcription termination element (*i.e.*, a sequence specifying mRNA 3'-end formation). The terms "transcription termination element," "polyadenylation site," "polyadenylation signal," and "sequence specifying mRNA 3'-end formation" are used interchangeably to refer to a site which defines the 3' end of a eukaryotic messenger RNA. The reporter gene is transcribed from a regulatable promoter in opposite orientation to the direction of retroviral transcription. The orientation of reporter gene transcription is significant, because it permits the introduction of transcription termination elements (such as might accompany a regulatory RNA sequence) downstream of the reporter gene in a way that does not interfere with production of vector-specific messenger RNAs. This permits the screening of, for example, oligo dT-primed cDNA libraries from mammalian cells which, by their nature, will contain transcription termination elements.

Candidate regulatory RNA sequences are inserted either 5' (upstream, in the transcriptional sense) or 3' (downstream) of the reporter gene. Expression from the regulatable promoter is controlled by synthesis of transactivator and/or transrepressor proteins, encoded on one or more regulatory vectors. The expression pattern of the reporter gene in the absence and presence of candidate regulatory RNA sequences is evaluated for level of expression and/or localization of the gene product. Sequences having regulatory activity are identified by such effects, and may then be isolated and characterized.

The invention further provides methods and compositions for identifying and characterizing molecules that interact with regulatory RNA sequences.

The invention additionally provides compositions and methods for regulating messenger RNA metabolism and expression, utilizing regulatory RNA sequences identified by the practice of the invention, in operative linkage with a therapeutic gene.

The invention also provides methods and compositions for modulating the expression of genes associated with pathogenesis. Such modulation can take the form of inhibition or activation of gene expression. Inhibition occurs at the post-transcriptional level through, for example, overexpression of a regulatory RNA sequence in a pathogenic cell so as to sequester factors which enhance or reduce the stability, translation, processing and/or localization of the mRNA product of a gene related to pathogenesis, thereby modulating the turnover, and/or the translation rate of the mRNA.

The invention also provides methods and compositions for expressing RNA molecules forming structures able to alter cellular or protein functions. Such RNAs can be expressed as a component of a mRNA molecule, using the methods of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Vector system to study the function of highly conserved regions in 3'UTRs

Figure 1A is a schematic representation of exemplary mammalian 3'UTRs studied here. Mammalian sequences were compared to their avian orthologues (except HuD, which was compared to *X. laevis*). Highly conserved regions (HCRs) are indicated by shaded boxes. Fragments that were cloned into the reporter vector are indicated by arrows. pA indicates that the endogenous polyadenylation site is present in the cloned fragment. The table in the right corner lists the species from which the HCRs were amplified, with the Genebank accession number and the HCR Database (ACUTS) identification number. Regions from which the primers were derived are indicated relative to the stop codon (+1 is the first base after the stop codon).

Figure 1B shows a schematic diagram of the reporter retrovirus designed to study the role of regulatory RNA sequences in post-transcriptional gene regulation. GUS: bacterial beta-glucuronidase. pA: bovine growth hormone polyadenylation signal. CMVm O7: tetracycline operator fused to the CMV minimal promoter. SV40PURO: puromycin resistance gene under the control of the SV40 promoter. SIN: self-inactivating retroviral LTR.

Figure 2. Effect of known HCRs on the expression of the GUS reporter

Figure 2A shows a FACS analysis of the sorted populations. Cells infected with low amounts of retroviruses (1 copy of provirus per cell) were loaded with FDGluc and sorted for positive GUS activity.

Figure 2B shows colorimetric detection of GUS activity. Cells from the sorted populations (see Figure 2A) were plated on tissue culture dishes and grown until they reached 50% confluence. They were then fixed for color detection.

Figure 2C shows a Northern blot analysis of steady-state mRNA levels. A Northern blot of total RNA (10 μg) from each of the HCR cell populations was simultaneously probed with two digoxigenin-labeled RNA probes directed against the GUS transcript (GUS) and against the ribosomal protein L32 (rpL) mRNAs. The

concentration of the rpL probe is limited in order to diminish the rpL signal, whose intensity at saturating concentration would obscure the quantification of the GUS signal.

Figure 3. Analysis of multicopy integration events

The upper panel shows a Northern blot (10 μg of total mRNA isolated from each of the HCR cell populations) probed simultaneously with three digoxigenin-labeled RNA probes against GUS, rpL and puro transcripts. The concentrations of both rpL and puro probes are limited to avoid obscuring the GUS signals. The signal intensity of GUS transcripts reflects the steady state levels of GUS mRNAs in the different HCR populations. This amount of mRNA is not only dependent on the differential stability of the transcripts but also on the amount of copies of the gene in the cell population. The lower panel shows graphically the steady state mRNA levels corrected for the amount of copies of the retroviruses in the different populations and relative to (-)HCR control cells. Quantitation of the signals was performed with the LumiAnalyst™ software. The bars were obtained by dividing the GUS values by the puro values and then by normalizing these ratios to the (-)HCR data. The rounded up corrected value is indicated on top of each bar. Data are representative of two experiments.

Figure 4. Effects on translation efficiency

In Figure 4A, the upper panel displays the steady state mRNA levels of the different populations relative to the (-)HCR population. Values are means ± standard errors from three experiments. The middle panel shows levels of GUS enzymatic activity normalized for protein levels. Values are means ± standard errors from three experiments. The lower panel displays the amount of protein produced per amount of mRNA. These values are derived from the ratio of protein levels to mRNA level. Values are ratio of the specific means ± standard errors of the ratio.

Figure 4B shows the absence of an effect on efficiency of translation by either the first or the second half of the vimentin 3'-HCR. Values are means ± standard errors from three experiments.

Figure 5. Stimulation by serum

Cells that had been incubated for 24 h in 5% horse serum (poor serum conditions) were stimulated by an incubation for 48 h in 20% serum (rich serum conditions: 15% CS.+ 5% FBS and compared to cells that remained in poor serum conditions. GUS enzymatic activities were determined and normalized for protein levels. The activity values relative to the (-)HCR counterpart are shown graphically: white bars indicate poor serum

conditions, grey bars indicate rich serum conditions; values are means ± standard error from three experiments.

Figure 6. Effect of hypoxia

Cells were incubated for 15 h in hypoxic conditions (4 ppm O_2) and compared to cells grown under normal conditions (*i.e.*, approximately 21% O_2, 5% CO_2). GUS enzymatic activities were determined and normalized for protein levels. For each cell population the values obtained for the 15 h hypoxic time point were divided by the values obtained for the nonhypoxic conditions. Values are means ± standard error from four experiments.

Figure 7. Regulation of the inducible CMVm promoter.

This figure shows a Northern Blot analysis of GUS mRNA from the (-)HCR cell population transduced either with MFG-rtTA or MFG-tTR vector, in the presence or absence of Doxycycline. The schemes at the bottom of the figure depict the influence of the different tetR on the inducible CMVm promoter in the absence or presence of Dox.

Figure 8. Regulation of a growth-inhibitory gene with a regulatable vector

Cells populations containing four retroviral vectors were selected. The vectors encoded 1) a tTR, 2) a rtTA, 3) GFP and puro resistance, and 4) p16 and $CD8^+$. Cells were propagated in the absence of doxycycline (to prevent growth arrest by p16), then exposed to various concentrations of dox to induce p16 expression.

In Figure 8A, 3-5 x 10^4 cells were grown in 6 wells plates in the absence of dox, then for 3 days at the indicated concentration of Dox.

In Figure 8B, 1.0-1.5 x 10^5 cells were cultured for 5 days in 10 cm plates in the presence of 10μg/ml Dox. The culture medium was then replaced with culture medium lacking Dox. The same region of the plate was photographed after 28 and 51 hours (the arrow indicates a large cell that did not move during that time period.

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional techniques in organic chemistry, biochemistry, oligonucleotide synthesis and modification, nucleic acid hybridization, molecular biology, microbiology, genetics, recombinant DNA, and related fields as are within the skill of the art. These techniques are fully explained in the literature. See, for example, Maniatis *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Laboratory Press (1982);

Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, Second Edition, Cold Spring Harbor Laboratory Press (1989); and Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons (1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996).

A retroviral vector system, and methods for its use, are provided for genomic screening to identify and characterize novel regulatory RNA sequences. In one embodiment of the invention, regulatory RNA sequences are located in the untranslated regions of mRNAs. In a two vector system, one vector contains a reporter gene in operative linkage to (*i.e.*, that is controlled by) a regulatable promoter, while the other vector drives expression of a trans-acting factor (transactivator or transrepressor) that can bind to the promoter and thus control expression of the reporter gene. Three-vector systems are also possible, comprising a reporter vector, a regulatory vector expressing a transrepressor, and a regulatory vector expressing a transactivator. The term "operative linkage" is used herein to refer to the juxtapositioning of two sequences (*i.e.*, a gene and a regulatory sequence, a gene and a transcription termination element, a promoter and an operator) such that a regulatory function is exerted by one of the sequences on the other. The expression pattern of the reporter gene in the absence and presence of candidate regulatory RNA sequences is evaluated for one or more aspects of messenger RNA metabolism, such as level of expression and/or localization of the gene product. Sequences that play a role in, for example, stability, translation, processing, transport and/or localization of mRNAs are identified, and may then be isolated from the reporter vector, and characterized. Such regulatory RNA sequences can also have wider effects, on processes such as cellular physiology and protein activity, for example.

The subject retroviral system is useful for studying untranslated RNAs, because there are minimal extraneous sequences. This provides an advantage because RNA secondary structure can alter its function and stability. Furthermore, through the practice of the invention, conditions can be obtained in which transcriptional effects are constant between reporter vectors containing different candidate regulatory RNA sequences, facilitating the direct comparison of the post-transcriptional effects of different regulatory RNA sequences.

The use of retrovirus based vectors provides for a fast and efficient system that yields large populations of transfected cells, rather than a few clones. This makes the system more useful and accurate than transient or stable transfections. Unlike other

methods of DNA transfer, with retroviral vectors it is possible to purify cell populations in which the majority of cells contains a single copy of the reporter vector, enabling more accurate comparisons between different cell cultures. The expression of the reporter expression-unit in an antisense direction (with respect to vector transcription) allows the cloning of UTRs containing their natural polyadenylation sequences. Thus, a gene encoding a detectable marker can be replaced with another gene of interest together with its natural polyadenylation sequence, which may play a regulatory role in its expression.

Due to their capability for stringent regulation of expression, the subject vectors are useful for studying gene expression in general. For example, they can be used for the regulated expression of cDNAs encoding toxic molecules that could adversely affect cell growth or differentiation at low levels, which require tight control of transcription. Figure 8 provides an example of regulated expression of the p16 gene, a gene involved in growth inhibition. The vector systems of the invention provide for regulated expression and characterization of untranslated RNAs or transcripts lacking an open reading frame, *e.g.* H19, ribozymes, antisense RNAs, synthetic, non-naturally-occurring RNA molecules, *etc.*, that can directly or indirectly affect cell metabolism, cell senescence, cell proliferation, differentiation and/or protein activity. For example, certain synthetic RNA sequences are capable of interacting with catalytic sites on proteins and with other RNA molecules. Controlled expression of therapeutic genes is achieved for gene therapy. Regulatable and tissue-specific expression of RNAs, *e.g.* ribozymes and antisense-RNAs, allows specific interference with the transcription and post-transcriptional regulation of, for example, oncogenic gene products involved in cancer.

The regulatory RNA sequences identified by the practice of the invention will allow regulation of heterologous genes in response to a variety of environmental factors which include, but are not limited to, stress, temperature change, oxygen tension, hypoxia, pressure, injury, ischemia, concentration of a mitogen or other growth-regulatory molecule, cell-cycle regulators, angiogenic factors, drugs or small molecules, and hormone concentration. Mitogens can include, but are not limited to, molecules such as epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, nerve growth factor, insulin, insulin-like growth factor-I (IGF-I), insulin-like growth factor-II (IGF-II), vascular endothelial growth factor, and endothelin.

Structure of Vectors

The vectors can be based on any virus or plasmid that is active in mammalian cells including, but not limited to, adenoviruses, poxviruses, herpesviruses (*e.g.*, CMV, EBV) togaviruses (*e.g.*, Semliki Forest Virus, Sindbis virus), retroviruses, lentiviruses, *etc.* Vectors can also include transposable elements, such as yeast retrotransposons and Ty elements, *Drosophila* transposable elements such as *copia,* human and mammalian transposable elements, *etc.* In a preferred embodiment, both the regulatory and the reporter vectors are based on mammalian retroviruses. An advantage of retroviruses is that their integration into the host genome allows for their stable transmission through cell division.

In one embodiment, the regulatory vector comprises a gene encoding either a transactivating (TA), or transrepressing (TR) factor that acts to regulate the expression of the reporter gene. Expression of the TR or TA factor from the regulatory vector is typically under the control of the retroviral LTR (as described below), although other promoters active in mammalian cells may also be used. In a preferred embodiment, the degree of activation or repression can be modulated by adjusting the concentration of a compound in the cellular growth medium. For example, tetracycline (tet) can be used to modulate the activity of tet-dependent transrepressors and transactivators. A single regulatory vector can be used, to provide regulation either by activation or repression, or two regulatory vectors, one encoding a transrepressor and the other encoding an transactivator, can be combined to provide repression and activation in a single cell. The presence of transactivator and transrepressor in a single cell provides very sensitive modulation of expression of the reporter gene. This permits the use of growth-inhibitory genes in the reporter vector, whose unregulated expression might be toxic to the cell. Cells containing such genes can be propagated under conditions in which gene expression from the reporter vector is repressed. At the appropriate time, gene expression can be activated by, for example, a change in tet concentration in the medium, to study the effect of the growth-inhibitory gene. It is also possible to over-express a cloned gene in this system, by relieving repression and/or stimulating activation.

Regulation of transcriptional activation is the result of interaction between transcriptional activators bound to *cis*-regulatory elements, factors bound to basal transcriptional elements and the activity of transcriptional mediators, or coactivators. The absence or presence of any of these factors affect the level of transcription. Additionally,

factors may be present in an inactive form, where the factors are activated through chemical modification, particularly as the result of a cellular signaling mechanism. In some cases, signaling molecules are able to act directly to activate transcription. Any of these factors may be used as regulators in the subject vectors. By "transactivator," "transactivating factor," or "transcriptional activator" is meant a polypeptide or polypeptide domain that facilitates transcription from a promoter. Where the promoter is an inducible promoter, the transactivator activates transcription in response to a specific transcriptional signal or set of transcriptional signals.

In one embodiment of the subject invention, the regulatory vectors comprise the MFG retroviral vector containing a tetracycline-regulatable transactivator (reverse tet-transactivator, rtTA) or tetracycline-regulatable transrepressor (tet-transrepressor, tTR). rtTA contains the VP16 activation domain (Gossen *et al.* (1995) *Science* **268:**1766-1769), and tTR contains the KRAB repression-domain (Deuschle *et al.* (1995) *Mol. Cell. Biol.* **15:**1907-1914). The tet-transfactors that are constitutively expressed from the 5'LTR promoter of MFG can bind to a tet operator sequence juxtaposed to the CMV minimal promoter and thus induce or repress promoter activity, respectively. Binding of the tet-transfactors is controlled by tetracycline or a derivative, *e.g.* doxycycline.

"Tetracycline repressor protein," tetracycline repressor polypeptide," "tetR polypeptide," "tetR protein" and "tT" are used interchangeably herein to mean a polypeptide that exhibits both 1) specific binding to tetracycline and/or tetracycline derivatives; and 2) specific binding to tetracycline operator (tetO) sequences when the tetR polypeptide is not bound by tetracycline or a tetracycline analogue(s). "TetR polypeptide" is meant to include a naturally-occurring (i.e., native) tetR polypeptide sequence and functional derivatives thereof. A reverse tetracycline repressor (rtT) refers to a polypeptide that binds to a tet operator sequence in the presence of tetracycline or a tetracycline analogue. For the purposes of the present invention, the major function of the tet repressor is to bind to a tet operator sequence, rather than to repress transcription, although, under certain conditions, steric repression of transcription by tet repressor in eukaryotic cells is possible. Accordingly, either of these tet repressors (tT and rtT) can be fused to either a eukaryotic repression domain or to a eukaryotic activation domain to generate transactivators or transrepressors active in eukaryotic cells. In one embodiment of the present invention, the tet repressor (tT) is fused to a eukaryotic repression domain

(the KRAB domain) to form a tranrepressor (tTR); and the reverse tet repressor (rtT) is fused to the VP16 transcriptional activation domain to form a transactivator (rtTA).

Thus, in a preferred embodiment the invention comprises a system in which reporter gene transcription is regulated by a tet transcriptional transrepressor (tTR) and a reverse tet transcriptional transactivator (rtTA). These molecules function as dimers; hence the tTR and rtTA polypeptide monomers contain a dimerization domain, in addition to either a transcriptional repression domain or a transcriptional activation domain. This presents a potential problem, as the most commonly used tet transactivators and transrepressors of the prior art share the same dimerization-domain (TetR, class B), leading to the possibility that heterodimers containing a repression domain and an activation domain will be formed in cells expressing both a tTR and a rtTA. To circumvent this problem, a preferred transrepressor utilizes the dimerization-domain of the bacterial TetR protein of class G, while the transactivator uses the class B dimerization domain. Since that class B and class G dimerization domains do not interact with each other to form dimers, heterodimers containing an activation domain and a repression domain cannot form. This allows the co-introduction of transactivators and transrepressors into the same cell. Thus, very efficient shutoff of reporter gene expression, combined with high induction levels, can be obtained by using two regulatory vectors: one encoding a tTR and the other encoding a rtTA.

The tetracycline system is advantageous in that the effect of the transcript of interest can be studied by turning it on and off with a soluble drug without affecting the rest of the cell. By contrast classical ways of studying stability-control require treating the whole cell with chemicals such as actinomycin D and cycloheximide.

Other transcriptional regulators of interest include the lambda phage repressor/operator system, E. coli *trp* operator/repressor, E. coli *lac* operator/repressor, E. coli *ara* operator/activator, ecdysone regulatory sequences, and the like. Where a bacterial or plant specific operator or repressor is used, it will be fused to a transcriptional regulatory domain that is functional in mammaiian cells.

The reporter vector contains a reporter gene encoding a detectable marker in operative linkage with a eukaryotic promoter. The reporter gene can also be in operative linkage to a transcription termination element, such that transcription of the reporter gene is terminated at the transcription termination element. In eukaryotes, the 3'-end of most transcripts is specified by a polyadenylation signal. Accordingly, in one embodiment of

the invention, a polyadenylation signal is located downstream (in the transcriptional sense) of the reporter gene. A preferred, efficient polyadenylation signal is that normally associated with the bovine growth hormone gene. Friedrich *et al.* (1991) *Genes Dev.* **5**:1513-1523; and Pfarr *et al.* (1986) *DNA* **5**:115-122. Certain candidate regulatory RNA sequences may include a polyadenylation signal; in these cases, the presence of a transcriptional termination element in the vector is optional.

The reporter gene is transcribed from a regulatable promoter in opposite orientation to the direction of retroviral transcription. Candidate regulatory RNA sequences are inserted either upstream (to the 5' side) or downstream (to the 3' side) of the reporter gene, by way of one or more cloning sites, as appropriate. Particularly preferred is a vector in which multiple cloning sites (also known as "polylinkers") are inserted on the 5' and 3' sides of the reporter gene (*i.e.*, between the promoter and the reporter gene, and between the reporter gene and the transcription termination site, if present), so that candidate regulatory RNA sequences can be cloned either upstream or downstream of the reporter gene. Modification of the basal level of expression from the regulatable promoter can be controlled by transactivator and/or transrepressor proteins, encoded on the regulatory vectors. In certain embodiments, expression of the reporter gene is modulated at the transcriptional level through substances which affect the activity of the transactivators and/or transrepressors. The expression pattern of the reporter gene in the absence and presence of candidate regulatory RNA sequences is evaluated for level of expression, transport and/or localization of the mRNA and/or protein gene products. To verify the post-transcriptional effect of a regulatory RNA sequence, it can be cloned into a region of the vector where it will not be transcribed into an RNA colinear with that of the reporter gene.

The ability to modulate expression from the reporter vector will be critical for situations in which a regulatory RNA sequence has growth-inhibitory effects, as in the case of, for example, cell-cycle regulators and tumor suppressors. In these situations, cell populations can be expanded with reporter vector expression switched off. The effect and properties of the regulatory RNA sequence can be studied at any time during growth of the cells, by switching on reporter vector transcription. Control of expression can be achieved, for example, by adjustment of the concentration of tetracycline (or a tetracycline analogue, such as doxycycline) in the medium, for cells in which expression is controlled by tet repressor-containing transactivators and/or transrepressors.

In one embodiment of the invention, the reporter vector has the structure depicted in Figure 1. The reporter vector includes the bacterial beta-glucuronidase (GUS) gene driven by a tetracycline-responsive cytomegalovirus (CMV) minimal promoter. Advantages of the GUS gene include its relatively small size (approximately 1.8 kilobases) and the many assays that are available for detection of its gene product, including histochemical, biochemical, and fluorescence-activated cell sorting (FACS) assays. The tetracycline-sensitivity results from the presence of a heptamerized tetracycline operator sequence (O7) fused to the promoter. This O7 sequence allows for the binding of tetracycline-controlled regulatory factors (see below) which can induce or repress the O7-promoter. The GUS expression unit, which includes the O7CMV-promoter, the GUS-reporter, and a synthetic transcription-termination, *i.e.* a polyadenylation (p(A)) sequence, is placed in opposite (antisense) orientation to retroviral transcription. This allows 3'UTR sequences that contain p(A) sites to be cloned downstream of the GUS reporter. These p(A) sites would cause premature termination of retroviral transcription in the virus-producing cells if transcription of the GUS expression proceeded in the same direction as retroviral transcription initiated at the retroviral LTR promoter. In addition, the GUS vector contains a selectable marker, *e.g.* a puromycin resistance gene, the expression of which is controlled by an internal SV40 promoter, in the sense direction (with respect to retroviral LTR-initiated transcription). The presence of the puromycin-resistance gene allows for the selection of cells transduced with the GUS vector.

Other reporter genes of interest include the green fluorescent protein (GFP), beta-galactosidase (Mohler *et al., supra*), beta-lactamase (Zlokarnik *et al.* (1998) *Science* **279**:84-88), and the like. Beta-lactamase is a small protein whose substrate is cell-permeable without the requirement for osmotic shock. In addition, enzymatic signal amplification is possible with the beta-lactamase reporter. The GFP gene allows live capture of the kinetics of cellular distribution. GFP fluorescence is stable, species-independent, and can be monitored noninvasively in living cells. GFP fluorescence persists in formaldehyde-fixed cells, and GFP is well-suited for double-labeling experiments with other fluorescent markers. In addition to wild-type GFP, red-shifted GFP variants are available, including EGFP (GFPmut1; Cormack *et al.* (1996) Gene), GFP-S65T (Heim *et al.* (1995) *Nature* 373:663-664), and RSGFP (Delagrave *et al.* (1995) *Bio/Technology* **13**:151-154).

The reporter vector will frequently include a marker that allows for selection of cells into which the DNA has been integrated, as opposed to cells which have not integrated the vector. Various markers are known in the art, particularly antibiotic resistance markers, such as resistance to G418 (neomycin), hygromycin, puromycin, and the like. Alternatively, negative selection can be used, where the marker is the *HSV-tk* gene, which will make the cells sensitive to agents such as acyclovir and gancyclovir.

The vectors include retroviral sequences that are required for packaging, integration and expression of the inserted selectable marker genes. The packaged vector is single stranded RNA, while the integrated provirus is double stranded DNA. The vector is "defective" in its inability to encode viral proteins required for productive infection. Replication requires growth in a packaging cell line that provides the *gag*, *pol*, and *env* proteins necessary for completion of the infectious cycle. For packaging of a vector genome to occur, a full-length transcript of the viral genome, initiated at the left-hand LTR, must be produced. If a regulatory RNA sequence comprising a transcription termination element is cloned into the vector in the same transcriptional orientation as the LTR-initiated transcript, production of a full-length viral transcript will be blocked, with a resultant decrease or abolition of packaging efficiency. In the reporter vectors of the invention, transcription of the reporter gene is in the opposite orientation to that of vector genome transcription. Hence, any transcription termination elements introduced along with a candidate regulatory RNA sequence will also be in an orientation opposite to that of vector transcription and will not block production of full-length vector transcripts. Thus, packaging efficiency will not be adversely affected.

The vectors contain a ψ sequence, which permits packaging of the retroviral genome. The ψ sequence is the region of the retroviral genome downstream from the 5' LTR, extending into the *gag* coding region (Danos *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464). It has been found that there is no sharp delineation of functional sequences within this region, but that including more of the native sequence will provide for better packaging efficiency. It is preferred that the *gag* and ψ sequences be derived from the same retrovirus species, *e.g.* MMLV, ALV, *etc.*, and that the ψ be positioned immediately downstream of the 5' LTR.

The sequences at the 5' and 3' termini of the vector are long terminal repeats (LTRs). A number of LTR sequences are known in the art and may be used. These include the MMLV-LTR; HIV-LTR; AKR-LTR; FIV-LTR; ALV-LTR; *etc.* Specific

sequences may be accessed through public databases. Various modifications of the native LTR sequences are also known. The LTR consists of sequence elements known as U3, R and U5 in the following order: U3-R-U5. In general, the R and U5 regions are essential for promoter activity, while some small deletions or substitutions may be made in the U3 region without the loss of promoter activity. The packaged retroviral RNA genome is known to exclude the 5' U3 sequence, and the 3' U5 sequence, thereby extending from R to R.

Generally the regulatory vector will have a functional 5' LTR. The 5' LTR acts as a strong promoter, driving transcription of the regulatory gene after integration into a target cell genome. Alternatively, expression of a regulatory gene can be controlled by any promoter active in mammalian cells, such as are known to those of skill in the art. Exemplary promoters include, but are not limited to, SV40 early, SV40 late, adenovirus major late, cytomegalovirus, polyomavirus, and poxvirus promoters.

For the reporter vector, in a preferred embodiment of the invention, the promoter function of the LTR is inactivated. This is accomplished by a deletion of the U3 region in the 3'LTR, including the enhancer repeats and promoter, that is sufficient to inactivate promoter function. After integration into a target cell genome, there is a duplication of the 5' and 3' LTRs, resulting in a transcriptionally defective provirus, termed a "self-inactivating" (SIN) vector. A SIN vector can have one or more exogenous promoters instead of or in addition to the LTR. Transcription of the selectable marker can be driven by the LTR or, alternatively, can be controlled by an exogenous promoter such as those described *infra*.

One may insert the vector sequence into an appropriate episomal vector, *e.g.* plasmid, EBV episome, BAC, YAC, *etc.*, and manipulate the vector by restriction, insertion of the desired gene with appropriate transcriptional and translational initiation and termination regions, and then introduce the plasmid into an appropriate packaging host. The vector may be further modified to include functional entities that find use in the preparation of the construct, amplification, transformation of the host cell, *etc.* At each of the manipulations, one may grow the plasmid in an appropriate host, analyze the construct to ensure that the desired construct has been obtained, and then subject the construct to further manipulation. When completed, the plasmid or excised virus may then be introduced into the packaging host for packaging and isolation of virus particles for use in the genetic modification.

Recombinant helper-free retrovirus production depends on a cell line that produces the necessary viral proteins. This ensures that the virus has the capacity to infect only one target cell and is then incapable of producing or transmitting virus to other cell types. These viral proteins include those encoded by the *gag*, *pol* and *env* genes, where each gene may produce polyproteins that are further processed after translation. A number of suitable packaging cell lines are known in the art, see for example, Mulligan (1993) *Science* **260**:926-932; Pear *et al.* (1993) *Proc. Natl. Acad. Sci. USA* **90**:8392-8396; and Mann *et al.* (1983) *Cell* **33**:153-159. Packaging cells are transfected with the retroviral DNA of interest by any suitable method, *e.g.* electroporation, $CaPO_4$ precipitation, DEAE-dextran, liposomes, particle bombardment, *etc.* The packaged virus is then collected from the supernatant of the packaging cells.

The host cell specificity of the retrovirus is determined by the envelope protein, *env* (p120). The envelope protein is provided by the packaging cell line, and is not encoded in the vector itself. Envelope proteins are of at least three types, ecotropic, amphotropic and xenotropic. Retroviruses packaged with ecotropic envelope protein, *e.g.* MMLV, are capable of infecting most murine and rat cell types. Ecotropic packaging cell lines include BOSC23 (Pear *et al., supra*) and ΨNX-E. Retroviruses bearing amphotropic envelope protein, *e.g.* 4070A (Danos *et al, supra.*), are capable of infecting most mammalian cell types, including human, dog and mouse. Amphotropic packaging cell lines include PA12 (Miller *et al.* (1985) *Mol. Cell. Biol.* **5**:431-437); PA317 (Miller *et al.* (1986) *Mol. Cell. Biol.* **6**:2895-2902) GRIP (Danos *et al.* (1988) *Proc. Natl. Acad. Sci. USA* **85**:6460-6464); and ψNX-A. The Phoenix (ΨNX) vectors are available by contacting the following internet address http://www-leland.stanford.edu/group/nolan/. Retroviruses packaged with xenotropic envelope protein, *e.g.* AKR *env*, are capable of infecting most mammalian cell types, except murine cells. Preferred packaging cell lines are derivatives of human 293 cells (ATCC CRL 1573), which rapidly take up and express high levels of transiently introduced DNA. Useful methods for the concentration of virus from cell supernatants include calcium mediated precipitation (Morling *et al.* (1995) *Gene Therapy* **2**:504-508); dialysis; affinity chromatography; or chromatography on a nickel column.

Screening for Post-transcriptional Regulatory Elements

Candidate regulatory RNA sequences are inserted into cloning sites in the regulatory vector, either 5' or 3' to the reporter gene as appropriate. Multiple cloning sites

(*i.e.*, "polylinkers") are preferred. Functional genomics is possible by testing untranslated RNA sequences for function. The subject vectors are useful for screening single, defined sequences. In another embodiment of the invention, the vectors are used to screen multiple sequences from a library (*e.g.*, a cDNA library) or database. The sequence or sequences is inserted into the polylinker cloning site by conventional recombinant techniques. Because transcription of the reporter gene in the reporter vector proceeds in an orientation opposite to that of the LTR-initiated transcript, it is possible, in the vectors of the invention, to insert sequences from an oligo dT-primed cDNA library (which contain transcription termination elements) without adversely affecting packaging of the vectors by interfering with production of full-length retroviral transcripts. This is especially important for screening cDNA libraries, as oligo dT-primed cDNA will contain polyadenylation sites specifying transcription termination.

In a preferred aspect of the invention, regulatory RNA sequences are derived from untranslated regions (UTRs) of messenger RNA. However, the practices of the invention are equally applicable to translated sequences, nontranscribed genomic sequences and synthetic sequences, such as those produced in combinatorial libraries. There are a number of sources for UTRs to be screened. In general, mRNA and cDNA sources are preferred to chromosomal sequences, in order to decrease the amount of non-transcribed sequences present. The mRNA may be from fetal, neonatal, junior (≤ 12 y) or adult tissue, where the cells are of the desired type, *e.g.* myogenic, endothelial, fibroblast, epithelial, neuronal, mucosal, cutaneous, hematopoietic, keratinocytes, hepatocytes, adipocytes, chondrocytic, osteogenic, and the like. Methods for the synthesis of cDNA from mRNA are well-known in the art (*e.g.*, Sambrook *et al.*, *supra*; and Ausubel *et al.*, *supra*) and may include amplification with specific or non-specific primers, e.g. RT-PCR, 3'RACE, *etc.* Regulatory RNA sequences from introns can be obtained from nuclear pre-mRNA, *i.e.*, hnRNA.

Published databases and commercially available libraries may be used as a source of UTRs. For example, EST sequences are frequently generated by poly-A priming of mRNA, and are therefore a rich source of 3'UTR sequences. These sequences are available through Genbank dbEST, or libraries are available. A database for highly conserved 3'UTR sequences can also be accessed. Duret *et al.* (1994) *Nucleic Acids Res.* 22:2360-2365; and Duret *et al.* (1996) *Comput. Appl. Biosci.* 12:507-510.

One may employ a genomic or cDNA library, where the fragments present in the library are relatively small as compared to the entire gene, or even the cDNA , so as to exclude at least about 25%, more usually at least about 75% of the total cDNA other than the 3'UTR. Methods for construction of cDNA libraries are well-known to those of skill in the art. See, for example, Ausubel *et al., supra*; and Sambrook *et al., supra*. Shortened cDNA sequences can be achieved with oligo dT priming, polymerase chain reaction rescue, and the like. As mentioned *infra*, the method of the invention is particularly suitable for screening oligo dT-primed cDNA libraries, because the transcription termination elements present in such libraries are cloned in opposite transcriptional orientation to the retroviral genome transcript, and will thus not interfere with the production of full-length LTR-initiated transcripts required for packaging.

One may use a subtraction library to enrich for regulatory elements associated with the target cell type. Alternatively, where the gene is known, one can cleave at a site adjacent to the terminal codon, and perform a polymerase chain reaction, using, as primers, oligonucleotides which recognize the sequence of the terminal codon and polyA, so as to limit amplification to sequences in the 3'UTR, or the like. For introns or other untranslated sequences outside of the 3' untranslated region of a mRNA, genomic DNA may be required. Alternatively, in some transcripts, such as hnRNA and nuclear pre mRNA, one or more introns is retained. Such retained introns could be identified in the cDNA, once the gene is shown to have a regulatory activity that is unassociated with the expressed protein. Non-mRNA transcripts, such as those produced by RNA polymerases I and III, are also a source of candidate regulatory RNA elements.

After the candidate sequences are inserted into the reporter vector, the vector is screened for the effect of the inserted sequence on expression of the reporter gene. Modulation of mRNA metabolism, including expression (*e.g.*, translation rate), stability, localization, transport, processing or developmental regulation, for example, as compared to the reporter gene in the absence of the candidate regulatory sequence, is indicative of a regulatory function. Modulation of mRNA metabolism by a regulatory RNA sequence can be assayed under various environmental conditions, including but not limited to mitogen stimulation, growth factor stimulation, tissue damage, ischemia, infection (viral or bacterial), variations in oxygen tension (*e.g.*, hypoxia), presence of cell-cycle regulators, stresses such as temperature (*e.g.*, heat shock and cold shock) and pressure; variations in ionic strength (*e.g.*, hyper-and hypotonicity, presence of specific ions, such as Ca^{2+}) and

pH, presence of metal ions, differentiation factors, angiogenic factors, senescence factors, toxins, carcinogens, teratogens, metabolites, extracellular matrix components, drugs, small molecules, and hormones.

Regulatory RNA sequences can be found in natural association with, or can be placed into operative linkage with genes encoding polypeptides of therapeutic utility, including but not limited to oncogenes, cell cycle regulators, differentiation factors, regulators of cell metabolism, regulators of senescence, molecules involved in cell-cell interactions, and molecules involved in cell-matrix interactions. Additional polypeptides of therapeutic utility, whose expression can be controlled by regulatory RNA sequences, include hormones, oncogenes, tumor suppressors, antigens, metabolic enzymes, senescence regulators, cell death genes, growth regulators, structural components and membrane components.

Typically, an assay will include one or more control samples having reporter gene only. Assays can also include control samples in which the reporter gene is operatively linked to a UTR or regulatory RNA sequence whose regulatory properties have already been characterized. Controls can also be performed in which a candidate regulatory RNA sequence is inserted in the regulatory vector in a region where it will not be included in the transcript of the reporter gene. Assays of interest for quantitating expression levels are well-known in the art. These include analysis of mRNA levels and analysis of the gene product. A number of methods are available for quantitating mRNA levels to determine mRNA stability, for example. These include, but are not limited to, hybridization (*e.g.*, Northern blots, dot blots, *etc.*), RNase protection and quantitative polymerase chain reaction (PCR). RNA localization may be determined by *in situ* hybridization, for example.

Detection of the gene product, to determine the translation rate of a mRNA, for example, may utilize enzymatic assays, particularly those that produce a colored or otherwise detectable product, gel electrophoresis, detection of fluorescence, chemiluminescence, antibody binding, *etc.* For example, GFP may be directly quantitated by a spectrophotometer. It can also be used in analysis by fluorescence-activated cell sorting (FACS), or photographed through a microscope for localization of the protein product. Enzymatic assays and substrates for the detection of reporters including, but not limited to, beta-galactosidase, GUS, beta-lactamase, alkaline phosphatase, peroxidase,

luciferase, *etc.* by biochemical, histochemical or FACS analyses are well-known to those of skill in the art and are commercially available.

Antibodies specific for certain reporter genes may be used in screening immunoassays. For example, detection may utilize staining of cells or histological sections, performed in accordance with conventional methods. The antibodies of interest are added to the cell sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody may be labeled with radioisotopes, enzymes, fluorescers, or other labels for direct detection. Alternatively, a second stage antibody or reagent is used to amplify the signal. Such reagents are well known in the art. For example, the primary antibody may be conjugated to biotin, with horseradish peroxidase-conjugated avidin added as a second stage reagent. Final detection uses a substrate that undergoes a color change in the presence of the peroxidase. Fluorescent or chemilumiscent substrates may be used in conjunction with enzymatic labels. The absence or presence of antibody binding may be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, *etc.*

Characterization and Utility of Regulatory UTRs

In one embodiment of the invention, regulatory RNA sequences which are derived from mRNA untranslated regions (UTRs) are transcribed and are normally associated with at least one exon in the mRNA. In other embodiments, UTRs from non-mRNAs, such as RNA polymerase I and RNA polymerase III transcripts, are obtained. The UTRs are transcribed from genomic DNA and may be processed, either being removed from the transcription product, by splicing, or remaining and forming part of the messenger RNA, as in the case of the 3'UTR. UTRs can be obtained from the 5' untranslated region, an intron, or the 3' untranslated region. Of particular interest are the 3'UTRs of genes associated with specific cellular functions.

Particularly preferred are highly conserved regions (HCRs), which are regions which are highly conserved between different organisms. In a preferred embodiment, highly conserved regions (HCRs) are defined as sequences within 3'UTRs that have retained greater than 70% homology within stretches of 100 or more nucleotides between organisms that diverged more than 300 million years ago. In one embodiment, conservation is determined as described by Duret *et al.* (1993) *Nucleic Acids Res.* **21**:2315-

2322. Briefly, sequences are compared between different classes of vertebrates, such as, for example, mammalian, avian and reptilian (and related organisms). While sequences having greater than 70% homology are preferred, sequences with conservation of less than 70% may be used, for example 30% or *e.g.* 50%. Furthermore, it will be apparent, to those of skill in the art, that conserved sequence elements shorter than 100 nucleotides can also be determined by this or similar methods. The UTRs are primarily associated with genes that are translated, and may include domains of RNA genes, such as the telomerase primer, RNA associated with a ribosome, *e.g.* 5S RNA, tRNA, or the like. A UTR can comprise all or a part of the naturally occurring untranslated region associated with an exon. Some HCRs may be located within the 3' UTR.

Usually, a naturally occurring HCR will be at least about 100 nucleotides, more usually, at least about 150 nucleotides. The active portion may be 60 nucleotides or more, frequently at least about 75 nucleotides, and not more than about 1000 nucleotides. A minimal sequence may be defined which is a portion of the naturally occurring HCR, which can fulfill at least 25%, preferably 50%, and most preferably 75% of the activity of the naturally occurring HCR. The HCRs may be associated with regulation of the expression of a number of genes in *trans*.

The regulatory RNA sequences can be used in a variety of ways, particularly as agents for the modulation of cell differentiation and/or division, cell metabolism and physiology, protein function, *etc.*; using the regulatory RNA sequence or its complementary sequence. In addition, regulatory RNA sequences may be used in affinity binding assays for identifying molecules which interact with regulatory RNA sequences to mediate their effects.

Levels of expression of endogenous pathologically-relevant genes that are controlled by regulatory RNA sequences can be modulated using the methods and compositions of the invention. For example, expression of an endogenous gene, involved in a pathological condition, that is controlled by a positively-acting regulatory RNA sequence (*i.e.*, one which acts to increase levels of the gene product) can be modulated by using the vectors of the invention to overexpress the same positively-acting regulatory RNA sequence in affected cells. Factors which interact with the positively-acting regulatory RNA sequence to mediate its effect will be sequestered by the exogenous overexpressed copies of the regulatory RNA sequence, preventing them from interacting with the endogenous RNA and thus blocking the positive effect of the regulatory RNA

sequence. Alternatively, expression of a RNA sequence that is complementary to the positively-acting regulatory RNA sequence (*i.e.*, an antisense sequence) can block the positive effect of the endogenous regulatory RNA sequence by forming a duplex which prevents binding of factors which normally interact with the sequence to mediate its effect. These and similar strategies can be applied, for example, to cells expressing an oncogene, to reduce expression of the oncogene product, or to virally-infected cells, to lower the levels of key transcripts whose products are required for viral replication or cell transformation.

In similar fashion, overexpression of exogenous regulatory RNA sequences can be used to increase the levels of a therapeutic gene product that is normally negatively regulated by an endogenous negatively-acting regulatory RNA sequence.

By virtue of the various regulatory activities of regulatory RNA sequences, nucleic acid compositions, proteins and small synthetic molecules which serve as mimetics may find use in this invention. Thus, the range of molecular weights of compounds, may vary from about 0.2 to 5 kD for small synthetic or small naturally-occurring organic molecules (distinguished from naturally occurring polymeric molecules, such as proteins, nucleic acids and polysaccharides), to naturally occurring polymeric molecules which may range to as high as about 200 kD, *e.g.* IgM. Generally the oligopeptides and oligonucleotides will be at least about 1 kD and not greater than about 10 kD, while the larger polymeric molecules will usually be at least about 10 kD, more usually at least about 30 kD.

High-throughput screening for regulatory RNA sequences is facilitated by the modular structure of the reporter vector and by the ability to sort cell populations by FACS to select cells containing only a single retroviral insertion This reduces the possibility, when cloning cDNA libraries into retroviral vectors, of obtaining cells with multiple retroviral insertions, wherein reporter gene expression might be controlled by several different regulatory RNA sequences.

Exemplary applications of regulatory RNA sequences

The regulatory RNA sequences of the invention can be used to regulate the expression of therapeutic genes. For example, genes whose overexpression would be desirable in certain situations include genes encoding inhibitors of viral replication, genes encoding bacteriostatic and/or bacteriocidal products, genes involved in cell death or apoptosis, tumor suppressors, oncogenes, genes encoding hormones, genes encoding cell

cycle regulators, and genes involved in proliferation. Regulated expression of genes involved in, for example, cell cycle regulation, cell metabolism and senescence may be desirable in other situations and can be achieved through the practice of the present invention. The practice of the invention allows regulation of therapeutic genes in response to environmental factors such as, for example, mitogens, angiogenic factors, and hypoxia, which are particularly relevant to cancer treatment. A therapeutic gene or a gene of interest can be regulated, through the mediation of a regulatory RNA sequence, by additional environmental conditions including but not limited to mitogen stimulation, growth factor stimulation, tissue damage, ischemia, infection (viral or bacterial), variations in oxygen tension (*e.g.*, hypoxia), presence of cell-cycle regulators, stresses such as temperature (*e.g.*, heat shock and cold shock) and pressure; variations in ionic strength (*e.g.*, hyper-and hypotonicity, presence of specific ions, such as Ca^{2+}) and pH, presence of metal ions, differentiation factors, angiogenic factors, senescence factors, toxins, carcinogens, teratogens, metabolites, extracellular matrix components, drugs, small molecules, and hormones. In addition to their use in regulating expression of a mRNA to which they are operatively linked, HCRs can be used to regulate the expression of antisense RNA molecules which can exert either positive or negative effects, in *trans*, on a second RNA molecule (see *supra)*. Examples of molecules whose expression can be negatively or positively controlled by regulatory RNA sequences, either in cis or in trans, include, but are not limited to, oncogenes, cell cycle regulators, differentiation factors, regulators of cell metabolism, regulators of senescence, molecules involved in cell-cell interactions, and molecules involved in cell-matrix interactions.

In addition, the methods and compositions of the invention can be used to regulate expression of potentially toxic or growth-inhibitory genes. Such genes can be placed under the tight regulation obtained with the tTR and rtTA-expressing regulatory vectors described *infra*. An example is provided *supra*, wherein use of the vectors of the invention allows both tight repression of the growth-inhibitory effects of the p16 gene and dose-dependent induction of p16 expression.

Regulation of a mRNA by a regulatory RNA sequence can include, for example, regulation of its stability, localization, nucleocytoplasmic transport, and translation rate. Changes in the aforementioned properties with respect to state of development or differentiation can also be mediated by a regulatory RNA sequence. This has implications for certain aspects of *ex vivo* gene therapy, chemotherapy, radiation therapy and immune

therapy in which stem cells are removed from a patient prior to treatment and reinfused into the patient after treatment. Current methods suffer from the inability to prevent stem cells from differentiating during culture outside of the patient. Thus, a regulatory RNA sequence which increases the stability and/or the translation rate of a mRNA encoding an inhibitor of differentiation, for example, is useful in this respect.

Regulatory RNA sequences can be used for analysis of mRNA and/or cDNA populations in functional genomics applications. For example, an array of different HCRs can be hybridized to a population of cDNAs isolated from cells that have been subjected to a particular environmental condition. The hybridization profile of the array is compared to a profile obtained by hybridizing cDNA from cells prior to their exposure to the particular environmental condition under study. HCRs to which cDNAs hybridize after, but not before exposure to the particular environmental condition are candidates for regulatory RNA sequences that mediate the response to the particular environmental condition under study.

The following examples are set forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention, and are not intended to limit the scope of what is regarded as the invention. Efforts have been made to insure accuracy with respect to the numbers used (*e.g.* amounts, temperature, concentrations, *etc.*) but some experimental errors and deviations should be allowed for. Unless otherwise indicated, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees centigrade; and pressure is at or near atmospheric.

EXAMPLES

The following example examines the regulatory properties of highly conserved regions (HCRs) within 3'UTRs that have retained greater than 70% homology within stretches of 100 nucleotides over 30 million years. A retroviral vector system is used with a selectable marker that allows rapid delivery of 3'UTR-reporter constructs to populations of thousands of cells within one to two weeks, avoiding problems associated with clonal analysis and long-term selection. Moreover, this vector is modular, thereby permitting direct comparison of different HCRs on gene expression, independent of 5'UTRs, promoters, protein coding regions, and polyadenylation signals. Since the 5'UTR, promoter, coding region and polyadenylation signal are constant in this vector, the

influence of HCRs on gene expression can be directly compared. Ten HCRs (c-fos, c-myc, transferrin receptor, bcl2, EF1α, vimentin, ornithine decarboxylase, fibronectin, HuD, and Ran), all of which are associated with proteins with a role in growth control, were examined. Using this methods and compositions of the invention, it is shown that HCRs can cause marked changes in mRNA and protein accumulation under steady state conditions and in response to changes in the cell milieu typical of sites where tumors develop. Nine of ten HCRs were found to decrease mRNA stability, to different extents. Two HCRs altered mRNA translation under steady state conditions. Four HCRs mediated responses to changes in mitogen level by increasing reporter protein levels 2-fold; whereas two HCRs exhibited a 6-fold difference in their response to another environmental stress, hypoxia.

MATERIALS AND METHODS

Plasmid construction

Vectors were constructed by using standard cloning procedures. To construct the reporter retrovirus, the following basic components were used: a 300 base pair (bp) SacI/XhoI fragment of the beta-geo clone (Friedrich *et al., supra*) for the bovine growth hormone polyadenylation signal; a 1.0 kilobase pair (kbp) SalI/ClaI fragment of pBabe puro (Morgenstern *et al.* (1990) *Nucleic Acids Res.* **18**:3587-3596) for the SV40-puro cassette; a 1.0 kbp SspI/BamHI fragment of pBabe puro for the 5' LTR; a 1.8 kbp Sse8387I/EcoRI fragment of pGUSN358→S (Clonetech) for the *E. coli* GUS gene; a 3.0 kbp BamHI/SspI fragment of the Retrotet vector (Hofmann *et al., supra*) for the self-inactivating (SIN) 3'LTR and most of the retroviral backbone; and a 0.5 kbp XhoI/EcoRI fragment of the Retrotet vector (Hofmann *et al., supra*) containing the O7CMV inducible promoter (seven tet operators juxtaposed to a CMV minimal promoter). All internal elements between the two LTRs were cloned sequentially in a Bluescript plasmid (Stratagene) containing a long polylinker. The plasmid backbone was then replaced with a retroviral backbone by XhoI/BamHI digestion.

The HCRs were amplified from the genome of the indicated species (Figure 1), except for the Ran and EF1-α HCRs, which were amplified from EST clones. The positions of the amplification primers are shown in Figure 1. The vim a HCR (Figure 4) was amplified with the primers complementary to nucleotides 3-22 and 138-157. The vim b HCR (Figure 4) was amplified with primers complementary to nucleotides 138-157 and

273-292. Amplifications were performed using the Expand™ High Fidelity PCR system (Boehringer Mannheim). Amplified fragments were purified by gel electrophoresis and cloned directly in the SrfI site of a modified pCR-SCRIPT™ SK(+) phagemid (Stratagene). This vector was modified by replacing the BamHI-KpnI polylinker with an AscI site, resulting in destruction of the BamHI site. Amplified HCRs were sequenced and found to be identical with the published sequences of those HCRs, except for the HuD HCR, which was found to be deleted for its 3'-terminal 23 nucleotides. All the HCRs were cloned in the reporter retrovirus with the restriction enzymes AscI (5'end) and BstXI (3' end).

The transrepressor retrovirus (MFG-tTR) was constructed by cloning a NcoI/BamHI fragment containing tetR-KRAB (Deuschle et al., *supra*) into the NcoI and BamHI sites of the MFG retroviral backbone. Riviere *et al.* (1995) *Proc. Natl. Acad. Sci. USA* **92**:6733-6737. The transactivator retrovirus (MFG-rtTA) was constructed by cloning a fragment encoding tetR-VP16 (Gossen *et al.* (1992) *Proc Natl. Acad. Sci. USA* **89**:5547-5551) into the NcoI and BamHI sites of the MFG retroviral backbone.

Tissue culture

The mouse embryonic fibroblast cell line C3H10T1/2 (further indicated as 10T1/2 cells) was purchased from ATCC (CCL-226; batch F-11839). 10T1/2 cells were propagated in growth media (GM) consisting of DME (Irvine Scientific) with 20% serum (15% calf serum (CS) + 5% fetal bovine serum (FBS), both from Hyclone). The retroviral-packaging cell line Phoenix-E was obtained from Dr. Gary Nolan (http://www-leland.stanford.edu/group/nolan/). Phoenix-E cells were grown in DME with 10% FBS. All media were supplemented with glutamine and penicillin/streptomycin according to the manufacturers recommendations. Cells were grown at 5% CO_2.

Transcriptional activity from the tetracycline-regulated O7CMVm promoter was controlled by the tetracycline analogue Doxycycline hydrochloride (Sigma). Relevant concentrations are indicated in the figures.

When necessary cultures were selected in the presence of the drug puromycin dihydrochloride (Sigma) at a concentration of 2.0-2.5 μg/ml. For selection, cells were grown in the presence of the drug for at least three generations.

Production of retrovirus: transfections and transductions

Retroviral particles were produced using a transient transfection procedure. Briefly, a 50-70% confluent 60 mm-dish of Phoenix-E cells was transfected with about 3 μg of plasmid DNA and incubated for 12-16 hrs. Subsequently, the medium was replaced and the cells further incubated overnight. Next, the retrovirus-containing medium was harvested, and either used directly for infection of target cells or frozen at -80°C for future use. After harvest, fresh medium was added to the dish and it was re-incubated and re-harvested. The procedure was repeated 3-4 times before the packaging-cells were discarded. Transfections were performed using the lipofectamine procedure (Life Technologies) according to the supplier's recommendations. In this case each 60 mm dish received a mix of 27 μl lipid + 3 μg DNA.

For transductions, target cells received undiluted viral supernatant to which polybrene (Aldrich) was added at a final concentration of 8 μg/ ml. Cells were then returned to the incubator for about 10-15 min, after which the dish containing the transduced cells was centrifuged at 2500 rpm for 45-60 min in a Beckman GPR centrifuge. The medium in the dishes was then replaced and cells were returned to the incubator for further growth. Transduction efficiency could be assessed by FACS analysis (see below) and was commonly above 50%. If lower efficiencies were required, the virus-containing supernatant was diluted prior to transduction.

FACS Analysis and cell sorting

10T1/2 cells were trypsinized and collected by centrifugation. Cells were resuspended in 100-200 μl of phosphate buffered saline (PBS) supplemented with 5% FBS (PBS/FBS), at room temperature. Cells were loaded with substrate by mixing the cell suspension with an equal volume of FDGlcU (Molecular Probes) dissolved in water, exposing the cells at 50% tonicity. After incubation at room temperature for 3-5 min, 2 ml of ice-cold PBS/FBS was added to the loaded cells to restore isotonicity. Addition of ice-cold buffer prevents the leakage of substrates and products across the cell membrane. The cold PBS/FBS mix contained 1 μg/ml propidium iodide for fluorescent labeling of the dead cells. Cells were maintained on ice until they were subjected to FACS analysis and sorting, which were performed as previously described. Nolan *et al.* (1988). *Proc. Natl. Acad. Sci. USA* **85**:2603-2607.

Colorimetric staining for GUS activity

The presence of the GUS reporter-gene product, beta-glucuronidase, was assayed by first fixing cells in 4% paraformaldehyde/0.25% glutaraldehyde (Sigma) in 100 mM sodium phosphate buffer, pH 6.6, for 3 min at room temperature, followed by reaction with 1-2 mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid (Gold Biotechnology) in a solution of 10 mM EDTA, 0.5 mM $K_3Fe(CN)_6$, $K_4Fe(CN)_6$, and 0.5 ml Triton X-100 (New England Nuclear) in 100 mM sodium phosphate buffer at 37°C, for several hours to overnight. The glucuronic acid stock was prepared in dimethylformamide at 40 mM and stored at -20°C, in the dark.

Northern Blot Analysis

Total RNA was isolated using the RNeasy Kit (Qiagen) according to the manufacturer's instructions. The RNA was denatured, electrophoresed in a 0.8% agarose formaldehyde gel (Davis *et al.* (1986). Basics Methods in Molecular Biology, New-York: Elsevier), transferred overnight to Nytran 0.45 μm membranes (Schleicher and Schuell) and UV-crosslinked in a Stratalinker (Stratagene) at an intensity of 70 mJ. Membranes were prehybridized for at least 1h in HB (50% deionized formamide, 2X SSC, 2% blocking solution (Boehringer Mannheim), 0.1 % N-Lauroylsarcosine, 0.02% sodium dodecyl sulfate) and hybridized overnight at 68°C in HB (2.5 ml/100 cm^2) containing digoxigenin-labeled RNA probes. The filters were then washed twice at room temperature for 5 min each in 2X SSC/0.1% sodium dodecyl sulfate, followed by two washes of 20 min each at 68°C in 0.1X SSC/0.1% sodium dodecyl sulfate. Chemiluminescent detection with the substrate CDP* (Tropix) was performed according to the protocol provided by Boehringer Mannheim with minor modifications. Spicher *et al.* (1994) *Dev. Biol.* **164**:72-86. The blots were imaged for 5 min followed by a 30 min exposure with the Lumi-Imager™ and the signal was quantified with the LumiAnalyst™ software (Boehringer Mannheim).

Digoxigenin (DIG)-labeled riboprobes were synthesized and tested according to Boehringer Mannheim instructions provided with the RNA labeling kit. The Gus riboprobe (1.8 kb) was synthesized with T7 RNA polymerase and the rpL32 (Kaspar *et al.* (1992) *J. Biol. Chem.* **267**:508-514) riboprobe (0.5 kb) was synthesized with T3 RNA polymerase.

GUS activity assay

Cells were harvested by centrifugation at 1000 rpm and the cell pellet was lysed in lysis-buffer consisting of Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, pH 7.0) with 0.2% NP-40 (Sigma). For each lysate, a dilution series in Z-buffer was plated in a 96-well plate format with each well containing 100 µl of sample volume. Using a multichannel pipetter, an equal volume of chemiluminescent substrate (Glucuron diluted 1:100 in water, Tropix) was added to each well according to the recommendations of the supplier. The plate was then incubated at room temperature for 3-4 hrs. Subsequently, 100 ul of Light-Emission-Accelerator solution (Tropix) was added followed immediately by analysis of the 96-well plate with the Lumi-Imager™(Boehringer Mannheim). Total protein content in the extracts was determined using the Biorad total protein assay and a microplate reader (Biorad model 450).

RESULTS

Many genes contain Highly Conserved Regions (HCRs) in their 3'UTRs

A comparison of nucleotide sequences of orthologous genes from different classes of vertebrates led to the discovery of Highly Conserved Regions (HCRs) in the untranslated regions (UTRs) of mRNAs. These HCRs were defined by Duret *et al.* (1993), *supra*, as regions characterized by more than 70% homology over at least 100 nt between organisms that diverged more than 300 million years ago. This strong conservation contrasts with an approximately 30% similarity expected between related UTR sequences that are not subject to selective pressure. Duret *et al.*, (1993), *supra*. Moreover, in greater than 10% of the cases analyzed, the degree of sequence conservation within HCRs in the 3'-untranslated region is higher than that in the protein coding region of the gene. HCRs in the 3' UTRs were found in 30% of the 370 known genes that are orthologous between mammals and birds. This high degree of sequence conservation between evolutionarily distant species suggests that HCRs in 3'UTRs have a significant function.

To test this hypothesis, the role of ten different HCRs in post-transcriptional regulation of gene expression was analyzed (Figure 1A). Three of the HCRs tested were from 3'UTRs that had well-documented effects (c-myc, c-fos, transferrin receptor); these served as controls for the assay. The UTRs harboring the other seven HCRs selected for study had not previously been rigorously investigated for potential regulatory function. They were chosen because they not only exhibited a high degree of evolutionary

conservation but were also associated with encoded products with known functions in cell cycle control and differentiation. For example, bcl2 prevents apoptosis and provides a selective growth advantage to many cell types. The gene encoding ornithine decarboxylase (ODC) is amplified in certain tumors (Thomas *et al.* (1991) *Breast Cancer Res. Treat.* **19**:257-267) and ODC levels correlate with proliferative capacity. Extracellular matrix components, such as fibronectin, alter growth and differentiation of many cell types. Proteins involved in translation control, like the eucaryotic elongation factor EF-1α, have been found to be associated with cancer. Sun *et al.* (1997) *Cancer Res.* **57**:18-23. Ran is a GTPase that that has been implicated in numerous processes like nuclear-cytosolic trafficking or cell cycle progression. Ren *et al.* (1995) *Mol. Cell. Biol.* **15**:2117-2124. Levels of vimentin, like those of c-myc, increase in mitogen-stimulated cells. Ferrari *et al.* (1990) *Cancer Res.* **50**:1988-1991. HuD has been implicated in both *Drosophila* and human cell differentiation and development. Thus, each of the HCRs analyzed is associated with a mRNA that encodes a product with a role in growth, differentiation, or transformation. To explore the possibility that these HCRs function in determining levels of expression of their associated proteins by altering mRNA stability or translation, a method was developed for efficient and well-controlled delivery of HCRs to cells. Following delivery of HCRs to cells, their effects both on steady-state mRNA and protein levels were analyzed, as well as their potential to mediate responses to stresses such as changes in growth factors or oxygen concentration.

The retroviral system for analysis of 3'UTRs

To determine whether HCRs within 3'UTRs have a role in regulating gene expression, a retroviral vector was constructed that allows a rapid and efficient assessment of HCR function in populations of hundreds of thousands of cells (Figure 1B). Two transcription units were included: one encodes a reporter gene (bacterial beta-glucuronidase, GUS), the other encodes a selectable marker (the puromycin resistance gene product, designated puro). The modular design of the vector, which has convenient linker sequences and restriction sites, allows HCRs to be inserted as desired and their effects on a standard reporter mRNA or protein to be studied in the absence of their own 5'UTR, promoter or polyadenylation (polyA) sequence. For example, the reporter gene, GUS, is flanked by two polylinkers allowing ready introduction of any given 5' or 3' UTR sequence.

Another feature of the retroviral vector is that transcription from the CMV minimal promoter used in all constructs is antisense to transcription from the viral LTR. This feature allows the retroviral expression of a 3'UTR sequence with a nonviral polyadenylation signal, either its own or a well characterized bovine growth hormone (bGH) poly (A) sequence. See, for example, Pfarr *et al.* (1986) *DNA* **5**:115-122. In the sense orientation, inclusion of a polyadenylation signal within a candidate regulatory RNA sequence, such as a 3'UTR, would not be possible, as it would arrest 5'LTR-initiated transcription prior to the transcription stop signal located in the 3'LTR, thereby impairing replication and production of infectious virus. However, the ability of a vector to accommodate an endogenous polyadenylation signal is important because many HCRs are juxtaposed to polyadenylation signals and cannot easily be separated from them (see Figure 1). For those HCRs which are separated from the polyA tails of their mRNAs, the bGH polyadenylation signal is used.

A self-inactivating (SIN) sequence within the LTR is necessary in order to express CMV promoter-directed transcripts of the reporter gene in antisense. Following infection of the target cell, the SIN vector, which contains a deletion in the enhancer and promoter sequences of the 3'LTR, transfers this deletion to the 5'LTR, resulting in the transcriptional inactivation of the provirus (Hofmann *et al.*, *supra*). This inactivation is essential in order to avoid production of transcripts directed by the strong regulatory elements located in the 5'LTR of the wild-type provirus, which would contain antisense sequences with respect to the reporter gene transcript.

GUS was selected as the reporter gene because it has many substrates which allow a range of assays. For example, the distribution of cells expressing the GUS enzyme can be analyzed in a transduced population and live cells separated and isolated according to their GUS activity using the fluorescence activated cell sorter (FACS). Lorincz *et al.* (1996) *Cytometry* **24**:321-329. In addition, GUS activity can be histochemically monitored at the single cell level using a colorimetric substrate. Kyle *et al.* (1992) Beta-Glucuronidase (GUS) Assay in Animal Tissue. In "GUS Protocols," S. R. Gallagher, ed. (San Diego: Academic Press), pp. 189-204. Finally, the amount of GUS protein produced can be determined by a highly sensitive quantitative chemiluminescent enzyme assay (commercially available from Tropix). A further advantage of GUS is that endogenous GUS activity in most mammalian cells (and therefore background levels of the enzyme) are low, especially when detection reactions are performed at neutral pH, which is optimal

for the bacterial but not the mammalian enzyme. Gallie *et al.* (1992). GUS as a Useful Reporter Gene in Animal Cells. In "GUS Protocols," S. R. Gallagher, ed. (San Diego: Academic Press), pp. 181-188. Moreover, the GUS coding sequence is relatively small (1.8 kb) which is advantageous in order that the other components of the retrovirus can be included.

Characterization and validation of an assay for regulatory RNA sequences

To characterize the retroviral vector system shown in Figure 1B, three HCRs from UTRs well known to alter mRNA stability (and therefore protein accumulation) were tested. In each vector tested, only the HCR differed. Mouse embryonic fibroblasts C3H10T1/2 (10T1/2) were transduced with retroviruses containing either no HCR or HCRs from the c-myc, c-fos or transferrin receptor (TfR) 3'UTRs (Figure 1A). Populations of thousands of cells obtained one week after transduction were analyzed for GUS expression by FACS. To control for potential differences due to copy number, these initial studies were carried out with populations in which the majority of cells harbored a single integrant. This was achieved by transducing cell populations at different viral titers and using only those populations in which less than 20% of the cells expressed GUS, as determined by FACS. According to the Poisson distribution, more than 96% of the GUS-expressing cells will have only one copy of the retrovirus. For each of the three HCRs tested, the sub-population of transduced cells expressing GUS above background levels was collected, expanded in culture, and reanalyzed by FACS. The enrichment was successful, as the FACS plots reveal that essentially all of the cells analyzed expressed significant GUS activity. The shape of the plots, or range of GUS expression, was similar for each of the HCR-expressing cell populations. This range reflects the random integration of the retrovirus in regions of the genome that differ in their transcriptional activity.

For each of these three HCR sequences, the peak, or mean value of GUS activity was determined for the cell population (Figure 2A). In this figure, the X-axis depicts GUS activity on a logarithmic scale and the shift of the peaks relative to the (-)HCR controls is due to the effect of the specific HCR sequences on the post-transcriptional regulation of expression of the GUS gene in the different cell populations. The c-fos HCR had the most marked effect, followed by c-myc and TfR. The mean value of the (-) HCR control cell population was 73, compared to 54 for the TfR HCR, 45 for the c-myc HCR, and 34 for

the c-fos HCR. These data show that FACS analysis is sufficiently sensitive to provide a rapid qualitative indication of the effects of a given HCR on protein expression levels in large populations of cells with random integration sites, and provides a high-throughput method of screening for regulatory RNA sequences.

A single cell analysis of GUS activity by histochemical assay gave results that paralleled those obtained by FACS, revealing a general range of activity in the cell population that was specific for each HCR. The c-fos HCR population had barely detectable GUS activity in this assay, followed by increasing amounts of blue staining in the c-myc HCR, TfR HCR and (-)HCR cell populations, respectively (Figure 2B), corroborating the results of the FACS analyses. The range in GUS expression exhibited among individual cells of a given HCR-expressing population provided further evidence that the cell populations containing a single copy of the integrated retrovirus are polyclonal. Such heterogeneity of expression levels also indicate that the effect of an HCR on the post-transcriptional regulation of GUS expression cannot be accurately assessed using cells derived from one or a few clones of stable integrants. These results underscore the need to study a polyclonal population of cells with a broad range of integration sites.

To analyze the effect of the HCRs on mRNA stability, the steady state level of expression of GUS mRNA in the three HCR expressing cell populations was determined by Northern Blot (Figure 2C). The blot was hybridized simultaneously with two different digoxigenin-labeled RNA probes that detect the mRNA encoding GUS and the mRNA encoding the ubiquitous ribosomal protein L32 (rpL), which serves as an internal control and allows correction for RNA loading. In this assay, as in the previous two assays, the accumulation of GUS mRNA was most profoundly altered by the c-fos HCR, followed by the c-myc and TfR HCRs, which also led to production of reduced levels of GUS transcripts relative to controls.

Thus, using three assays (FACS, histochemistry, and Northern), all three HCRs led to the production of significantly reduced amounts of reporter (GUS) protein and mRNA levels. These findings confirm the results obtained with the 3'UTRs of these genes by others. In addition, the assays described here allow a direct comparison of the effects of c-fos, c-myc and TfR HCRs on post-transcriptional gene regulation, since the only variable is the HCR. The results show that the HCR of c-fos reduces mRNA accumulation to a greater extent than that of c-myc or TfR. These findings validate the retroviral test system

for the study of post-transcriptional changes in gene expression mediated by regulatory RNA sequences.

Identification of novel HCRs that affect post-transcriptional gene expression

Seven additional HCRs were introduced into reporter vectors containing a GUS gene. To analyze the effect of these 7 "test" HCRs on mRNA stability, steady state levels of expression of GUS mRNA in the different HCR-expressing cell populations was determined by Northern Blot. 10T1/2 fibroblasts were transduced with retroviral vectors containing either no HCR, an HCR from one of the 3 well characterized 3'UTRs (c-fos, c-myc and TfR) or one of the 7 "test" HCRs. The test HCRs had not previously been characterized with respect to their function in post-transcriptional regulation. The 11 cell populations were sorted by FACS for GUS expression and expanded by growth in culture. To expedite and facilitate the analysis of several HCRs in parallel, cell populations were isolated irrespective of transduction efficiency.

In this experiment, the SV40-puro transcription unit was used to control for the number of integrants per cell in each cell population, precluding the need for a genomic Southern blot. Both GUS and puro transcription units are on the same vector, with the GUS transcription unit being potentially affected by the HCR, whereas expression of the puro transcription unit is not. Thus, expression of puro mRNA, as detected on Northern blots following correction for RNA loading (by comparison to rpL32 mRNA levels) is proportional to the number of retroviral integrants.

The Northern blot in Figure 3A was hybridized simultaneously with three digoxigenin-labeled RNA probes specific to transcripts for GUS, puro, and rpL32. Accurate quantitation of the differences in chemiluminescent signals shown in Figure 3 was made possible by using a highly sensitive luminometer (Lumi-Imager™; Boehringer Mannheim) that exhibits a linear dynamic range over 1:10000. This linear range, which is 100-fold greater than that obtained with X-ray films, allows quantitation of a wide range of very strong and very weak signals with a single exposure. To determine the relative stability of the GUS-HCR mRNAs in each cell population, the values obtained for GUS were divided by the values obtained for puro and expressed as a percentage of the values obtained for the (-)HCR population.

The hypothesis that GUS/puro signals provides an indication of the specific effects of the HCR was validated, as similar results were obtained for the three well-characterized

UTRs (c-myc, c-fos, and TfR). The results relative to the (-)HCR control for these three HCRs were similar for single copy integrants (Figure 2C) and for multicopy integrants (Figure 3). The destabilizing properties of the HCR of c-fos were most profound, with the c-myc and TfR HCRs also causing significant reductions in GUS mRNA accumulation. Thus, the results obtained either by using puro mRNA levels to normalize multicopy integrants or by selection of single copy integrants (Figure 2C) were similar, validating the simpler, less labor-intensive multicopy approach used in all subsequent experiments.

The results shown in Figure 3 demonstrate the striking effects of HCRs on mRNA accumulation. With the exception of the Ran HCR, all of the HCRs tested had a destabilizing effect on GUS mRNA. The Ran HCR yielded mRNA steady state levels comparable to the (-)HCR control population. The HCR of c-fos was the most potent destabilizer, leading to an accumulation of less than 10% of the levels of control (-)HCR transcripts, a level readily quantitated by the Lumi-Imager™, but not revealed using the exposure parameters of X-ray film (Figure 3). The second most destabilizing HCR was that of ornithine decarboxylase (ODC), which resulted in 30% of the levels of control (-) HCR transcripts. This effect of the ODC 3'UTR on mRNA stability has not previously been reported. The remaining 5 test HCRs (EF1-α, vimentin, fibronectin, bcl2, and HuD) all caused a 30 to 50% decrease in steady state GUS mRNA levels.

Efficiency of translation is altered by HCRs

To determine the effect of HCRs on translation, the steady state levels of GUS mRNA (Figure 4A) were compared to the steady state levels of GUS enzyme activity (Figure 4B) for a number of HCR-expressing cell populations. For this purpose, GUS enzyme activity was determined by a chemiluminescent assay performed in 96 well plates and quantified using the Lumi-Imager™. As a measure of translation, GUS activity values were corrected for the amount of total protein in the extracts and expressed relative to the values obtained for the (-)HCR population.

Most of the HCRs analyzed had no effect on the translation efficiency of GUS transcripts. This was clear from a comparison of the top panel, GUS mRNA accumulation, with the middle panel, GUS activity (*i.e.*, protein accumulation). In most cases, the amount of mRNA determines the amount of protein, as depicted in the bottom panel of Figure 4A, which shows the ratio of GUS activity to GUS mRNA. Two HCRs, however, led to a marked lack of correlation between mRNA and protein levels, suggesting that they

have an important effect on GUS mRNA translation. The c-fos HCR led to a decrease not only in mRNA accumulation but also in protein accumulation; the net effect being a 3-fold translational repression of the GUS transcript. By contrast, the vimentin HCR enhanced translation by 2-fold. These results demonstrate that specific HCRs can alter gene expression post-transcriptionally at the level of mRNA translation.

Since the vimentin HCR (Figure 1A) is comprised of two highly conserved regions (80-90% conservation) separated by a region that is not as highly conserved, a test was conducted to determine if one of these two regions would suffice for the observed enhancement of translation. These two regions were therefore separately cloned and the effect of each sub-region was assayed. The GUS Activity/GUS mRNA ratio was individually determined for each of the two sub-regions in transduced cell populations. In neither of these cases was the strong effect obtained with the intact vimentin HCR detected. Rather, the activity/mRNA ratios were similar to those obtained with the (-)HCR control population. These results suggest that the vimentin HCR requires both elements and cannot be disrupted without altering its function.

HCRs mediate responses to changes in mitogen concentration

The ability of HCRs to alter gene expression at the post-transcriptional level in response to a stress, such as a change in growth factor concentration in the culture medium, was tested. For this purpose, HCR-transduced cell populations were grown in poor serum media for 14 hours. Thereafter, half of the cultures remained in poor serum media while the other half were shifted to rich serum media. GUS activity was determined in the different HCR-expressing cell populations 24 hours later (Figure 5). Five of the HCR-expressing cell populations showed no difference in GUS activity in response to changes in mitogen levels, including c-fos, TfR, bcl2, EF1α, and vimentin. By contrast, four of the HCRs responded to mitogen stimulation by inducing a two-fold increase in GUS protein levels. Thus, certain HCRs mediate post-transcriptional mechanisms by which cells can significantly increase the levels of specific growth related proteins.

HCRs can respond to changes in oxygen concentration

Since regions of tumors are often hypoxic, and cells that are transformed adapt to such changes and continue to grow, the ability of HCRs to alter gene expression in response to the stress induced by changes in oxygen tension was tested (Figure 6).

Populations of cells expressing either the c-fos HCR or the bcl2 HCR were cultured at 4 parts per million O_2 for 15 hours. GUS activity was determined and expressed as a function of the GUS activity in control cultures grown in parallel under normal culture conditions of 21 % O_2 and 5% $C0_2$. Under conditions of hypoxia, the c-fos HCR induced an increase in GUS protein levels, and the bcl2 HCR reduced GUS protein levels, such that the c-fos HCR produced GUS levels that were 6-fold greater than those observed for the bcl2 HCR. Thus, in response to an environmental stress such as low oxygen, HCRs can have marked post-transcriptional effects on gene expression that may be critical for cell survival.

Modulation of HCR expression levels

The utility of this retroviral system for the study of HCR function can be enhanced by providing for modulation of the dosage of the HCR. In theory, this should be possible with the reporter vector described herein, since seven copies of the tet-operator precede the CMV minimal promoter (Figure 1). All of the experiments presented thus far examined the effects of HCRs on gene expression using basal expression from the CMV minimal promoter. To test whether the expression of the promoter could be regulated, cells containing the a reporter vector without an HCR were super-infected with a second regulatory vector. The regulatory retrovirus can encode either a tetracycline-regulatable fusion protein comprised of the *E. coli* tet repressor (that binds the tet operator in the absence of tet or its analogues) fused to the KRAB repression domain of the human Kox1 zinc finger protein (tTR, Deuschle et al., *supra*) or it can encode a mutant ("reverse") tet repressor (which binds tet operators in the presence of tet or its analogue) fused to the viral VP16 activation domain (rtTA). Gossen *et al.* (1995) *supra.* In the absence of tet (or its analogue dox), tTR binds the tet operator and represses transcription. In the presence of dox, the tTR is sterically inhibited from binding the tet operator adjacent to the CMV minimal promoter, so transcription is not repressed and proceeds at a basal level. On the other hand, when the (-)HCR cell line is infected with the regulator retrovirus encoding the mutant ("reverse") tet transactivator, basal transcription levels are obtained in the absence of dox, while, in the presence of dox, the rtTA binds the tet operator, stimulating maximal induced levels of transcription. Figure 7 shows that the basal level of transcription used in all of the studies reported thus far can be decreased or increased at will in the same HCR-expressing cell population by altering the concentration of dox. An even wider range of

dosages can be obtained by using two regulatory vectors: one encoding a tTR and the other encoding a rtTA. Thus, by converting the system to a binary or ternary retroviral vector system, the effects of dosage on cell proliferation and differentiation can be readily assayed.

Regulated expression of a growth-inhibitory gene

The p16 gene induces cell-cycle arrest by inhibiting the cyclin-dependent kinases CDK4 and CDK6. This prevents them from phosphorylating and inactivating the tumor suppressor gene RB, thereby inhibiting progression through the cell cycle. Thus, cells expressing high levels of p16 fail to divide.

An experiment was designed to test whether expression of p16 could be regulated tightly enough to prevent its growth-inhibitory activities from being expressed in transfected cells. C3H10T1/2 cells were transduced with a reporter retrovirus encoding GFP as an inducible reporter and puromycin resistance as selectable marker. This GFP-expressing cell line was superinfected with two regulator retroviruses: one expressing rtTA and the other expressing tTR. The combination of the activator and repressor in the same cell line enables tight control of reporter gene expression. This inducible cell line was then superinfected with a reporter retrovirus encoding the mouse p16 gene as an inducible cassette and the gene coding for CD8 as selectable marker. CD8-positive cells were selected by cell sorting. The CD8- and GFP-positive cell population was propagated in culture medium lacking doxycycline (dox), to prevent cell growth arrest through p16 expression.

Cells were then exposed, for three days, to different concentrations of dox and photographed. The results show a dose-response of the cells to increased p16 expression. As seen in Figure 8A, with increasing concentrations of dox (which leads to increased p16 expression) fewer cells are observed. The effect of p16 on cell growth is reversible, upon removal of dox from the culture medium. Figure 8B shows that, after 5 days culture at 10μg/ml Dox, withdrawal of dox led to an increase in cell density. Thus, the regulatable vector systems of the invention can be used to hold in check the expression of potentially toxic genes or regulatory RNA sequences, and limit their expression to particular times or locations by simple adjustments of the cellular growth medium.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were

specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope.

CLAIMS

What is claimed is:

1. A viral vector for identification of a regulatory RNA sequence, the vector comprising viral regulatory sequences and a reporter gene, wherein the reporter gene is in operative linkage to a promoter, wherein a cloning site is located downstream of the reporter gene, and further wherein transcription of the reporter gene proceeds in an orientation opposite to that of transcription initiated at the viral regulatory element.

2. The vector according to claim 1, wherein the viral vector is a retroviral vector.

3. The vector according to claim 2, wherein the viral regulatory sequences comprise long terminal repeat (LTR) sequences.

4. The vector according to claim 3, wherein, the LTR sequences are self-inactivating.

5. The vector according to claim 1, wherein the vector further comprises a polyadenylation signal, wherein the downstream cloning site lies between the reporter gene and the polyadenylation signal.

6. The vector according to claim 5, wherein the polyadenylation signal is the bovine growth hormone polyadenylation signal.

7. The vector according to claim 1, wherein the vector further comprises a cloning site between the promoter and the reporter gene.

8. The vector according to claim 4, wherein the vector further comprises a cloning site between the promoter and the reporter gene.

9. The vector according to claim 8, wherein the promoter is a regulatable promoter.

10. The vector according to claim 9, wherein the promoter is a cytomegalovirus minimal promoter.

11. The vector according to claim 10, wherein the promoter is operatively linked to one or more *tet* operator sequences.

12. The vector according to claim 1, wherein the vector further comprises a gene encoding a selectable marker.

13. The vector according to claim 12, wherein the selectable marker encodes antibiotic resistance.

14. The vector according to claim 13, wherein the selectable marker encodes resistance to puromycin.

15. The vector according to claim 11, wherein the vector further comprises a gene encoding a selectable marker.

16. The vector according to claim 15, wherein the selectable marker encodes antibiotic resistance.

17. The vector according to claim 16, wherein the selectable marker encodes resistance to puromycin.

18. The vector according to claim 1 wherein the regulatory RNA sequence is derived from the 3'-untranslated region of a messenger RNA.

19. The vector according to claim 18 wherein the regulatory RNA sequence comprises a highly conserved region.

20. The vector according to claim 1 wherein the regulatory RNA sequence mediates a response to an environmental condition selected from the group consisting of stress, temperature, oxygen tension, growth factor concentration, ionic strength, pH, tissue damage, ischemia, infection, concentration of cell-cycle regulators, pressure; metal ion concentration, differentiation factors, angiogenic factors, senescence factors, toxins, carcinogens, teratogens, metabolites, extracellular matrix components, drugs, small molecules, and hormones.

21. The vector according to claim 20, wherein the growth factor is selected from the group consisting of epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, nerve growth factor, insulin, IGF-I, IGF-II, vascular endothelial growth factor, and endothelin.

22. The vector according to claim 20, wherein the regulatory RNA sequence mediates a response to hypoxia.

23. The vector according to claim 15, wherein the regulatory RNA sequence regulates a process selected from the group consisting of RNA localization, RNA stability, messenger RNA translation, nucleocytoplasmic transport of RNA and RNA processing.

24. The vector according to claim 15, wherein the regulatory RNA sequence is involved in developmental regulation of processes selected from the group consisting of localization, translation, stability, processing and nucleocytoplasmic transport.

25. A viral vector for regulation of an RNA molecule, the vector comprising a viral regulatory element and a nucleotide sequence encoding the RNA molecule, wherein the nucleotide sequence encoding the RNA molecule is in operative linkage to a promoter and to a regulatory RNA sequence, and further wherein transcription of the nucleotide

sequence encoding the RNA molecule proceeds in an orientation opposite to that of transcription initiated at the viral regulatory element.

26. The vector according to claim 25 wherein the regulatory RNA sequence is derived from the 3'-untranslated region of a messenger RNA.

27. The vector according to claim 26 wherein the regulatory RNA sequence comprises a highly conserved region.

28. The vector according to claim 25 wherein the regulatory RNA sequence mediates a response to an environmental condition selected from the group consisting of stress, temperature, oxygen tension, growth factor concentration, ionic strength, pH, tissue damage, ischemia, infection, concentration of cell-cycle regulators, pressure; metal ion concentration, differentiation factors, angiogenic factors, senescence factors, toxins, carcinogens, teratogens, metabolites, extracellular matrix components, drugs, small molecules, and hormones.

29. The vector according to claim 28, wherein the growth factor is selected from the group consisting of epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, nerve growth factor, insulin, IGF-I, IGF-II, vascular endothelial growth factor, and endothelin.

30. The vector according to claim 28, wherein the regulatory RNA sequence mediates a response to hypoxia.

31. The vector according to claim 25, wherein the regulatory RNA sequence regulates a process selected from the group consisting of RNA localization, RNA stability, messenger RNA translation, nucleocytoplasmic transport of RNA and RNA processing.

32. The vector according to claim 25, wherein the regulatory RNA sequence is involved in developmental regulation of processes selected from the group consisting of localization, translation, stability, processing and nucleocytoplasmic transport.

33. The vector according to claim 25, wherein the RNA molecule is a messenger RNA.

34. The vector according to claim 33, wherein the messenger RNA encodes a polypeptide involved in cell proliferation.

35. The vector according to claim 34, wherein the messenger RNA encodes a polypeptide selected from the group consisting of oncogenes, cell cycle regulators, differentiation factors, regulators of cell metabolism, regulators of senescence, molecules involved in cell-cell interactions, and molecules involved in cell-matrix interactions.

36. The vector according to claim 335, wherein the messenger RNA encodes a polypeptide involved in cell death.

37. The vector according to claim 36, wherein the messenger RNA encodes a polypeptide selected from the group consisting of apoptotic genes and growth-inhibitory genes.

38. A method for identifying a regulatory RNA sequence utilizing the vector according to claim 1, the method comprising the steps of:

(a) inserting a candidate regulatory sequence into the cloning site to generate a regulatory construct,

(b) exposing a host cell population to the regulatory construct obtained in step (a),

(c) selecting a sub-population of stably transformed cells,

(d) determining expression of the reporter gene in the stably transformed sub-population,

(e) determining the expression of the reporter gene in a sub-population of cells that are stably transformed by the vector lacking the candidate regulatory sequence, and

(f) comparing expression of the reporter gene determined in steps (d) and (e).

39. A method for constructing a cDNA library in a retroviral vector, wherein the cDNA inserts comprise transcriptional termination elements, using the vector according to claim 1.

40. The vector according to claim 1, lacking a reporter gene and further comprising a regulatory RNA sequence.

41. A method for regulating expression of an RNA , the method comprising the steps of:

(a) placing a nucleotide sequence encoding the RNA into a vector according to claim 40, and

(b) introducing the vector of step (a) into a host cell under conditions wherein transcription of the nucleotide sequence occurs.

42. A method for regulating expression of an RNA , the method comprising the steps of:

(a) placing a nucleotide sequence encoding the RNA into the vector according to claim 40, and

(b) introducing the vector of step (a) into a host cell under conditions wherein transcription of the nucleotide sequence occurs;

wherein the nucleotide sequence encoding the RNA comprises a transcription termination element.

43. The method according to claim 41, wherein transcription is regulated by a transrepressor.

44. The method according to claim 41, wherein transcription is regulated by a transactivator.

45. The method according to claim 41, wherein transcription is regulated by a reverse transrepressor.

46. The method according to claim 41, wherein transcription is regulated by a reverse transactivator.

47. The method according to claim 41, wherein transcription is regulated by a combination of a transrepressor and a reverse transactivator.

48. The method according to claim 41, wherein transcription is regulated by a combination of a reverse transrepressor and a transactivator.

49. The method according to claim 41, wherein the regulatory RNA sequence regulates a process selected from the group consisting of messenger RNA stability, messenger RNA localization, messenger RNA translation and nucleocytoplasmic transport of messenger RNA.

50. The method according to claim 41, wherein the messenger RNA encodes a therapeutic polypeptide.

51. The method according to claim 50, wherein the therapeutic polypeptide is selected from the group consisting of hormones, oncogenes, tumor suppressors, antigens, metabolic enzymes, senescence regulators, cell death genes, growth regulators, structural components and membrane components.

52. A cell into which has been inserted the vector of claim 40.

53. A vector encoding a regulatory RNA sequence, wherein the regulatory RNA sequence is operatively linked to a regulatable promoter.

54. A cell into which has been inserted the vector of claim 53.

55. A method for regulating a mRNA, wherein said mRNA comprises a regulatory RNA sequence, the method comprising the following steps:

(a) introducing a vector according to claim 53 into a host cell in which the mRNA is expressed; and

(b) culturing the cell of (a) under conditions wherein the regulatory RNA sequence is overexpressed;

such that the overexpressed regulatory RNA sequence interacts with cellular factors, thereby preventing their interaction with the regulatory RNA sequence of the mRNA.

56. A vector encoding a sequence that is complementary to a regulatory RNA sequence, wherein the sequence that is complementary to a regulatory RNA sequence is operatively linked to a regulatable promoter.

57. A method for regulating a mRNA, wherein said mRNA comprises a regulatory RNA sequence, the method comprising the following steps:

(a) introducing a vector according to claim 56 into a host cell in which the mRNA is expressed; and

(b) culturing the cell of (a) under conditions wherein an antisense RNA regulatory sequence is overexpressed, wherein said antisense sequence is complementary to the regulatory RNA sequence present in the mRNA,

such that the overexpressed antisense sequence interacts with the regulatory RNA sequence present in the mRNA, thereby blocking the function of the regulatory RNA sequence present in the mRNA.

58. A method for identifying molecules which interact with a regulatory RNA sequence, the method comprising the following steps:

(a) introducing a vector according to claim 55 into a host cell,

(b) culturing the cells of (a) under conditions wherein the regulatory RNA sequence is overexpressed,

(c) isolating the regulatory RNA sequence from the cells, and

(d) identifying molecules associated with the isolated regulatory RNA sequence.

59. A molecule which interacts with a regulatory RNA sequence, said molecule identified by the method of claim 58.

60. A regulatory RNA sequence isolated by the method of claim 38.

61. The method according to claim 41, wherein the vector is introduced into the host cell *in vivo*.

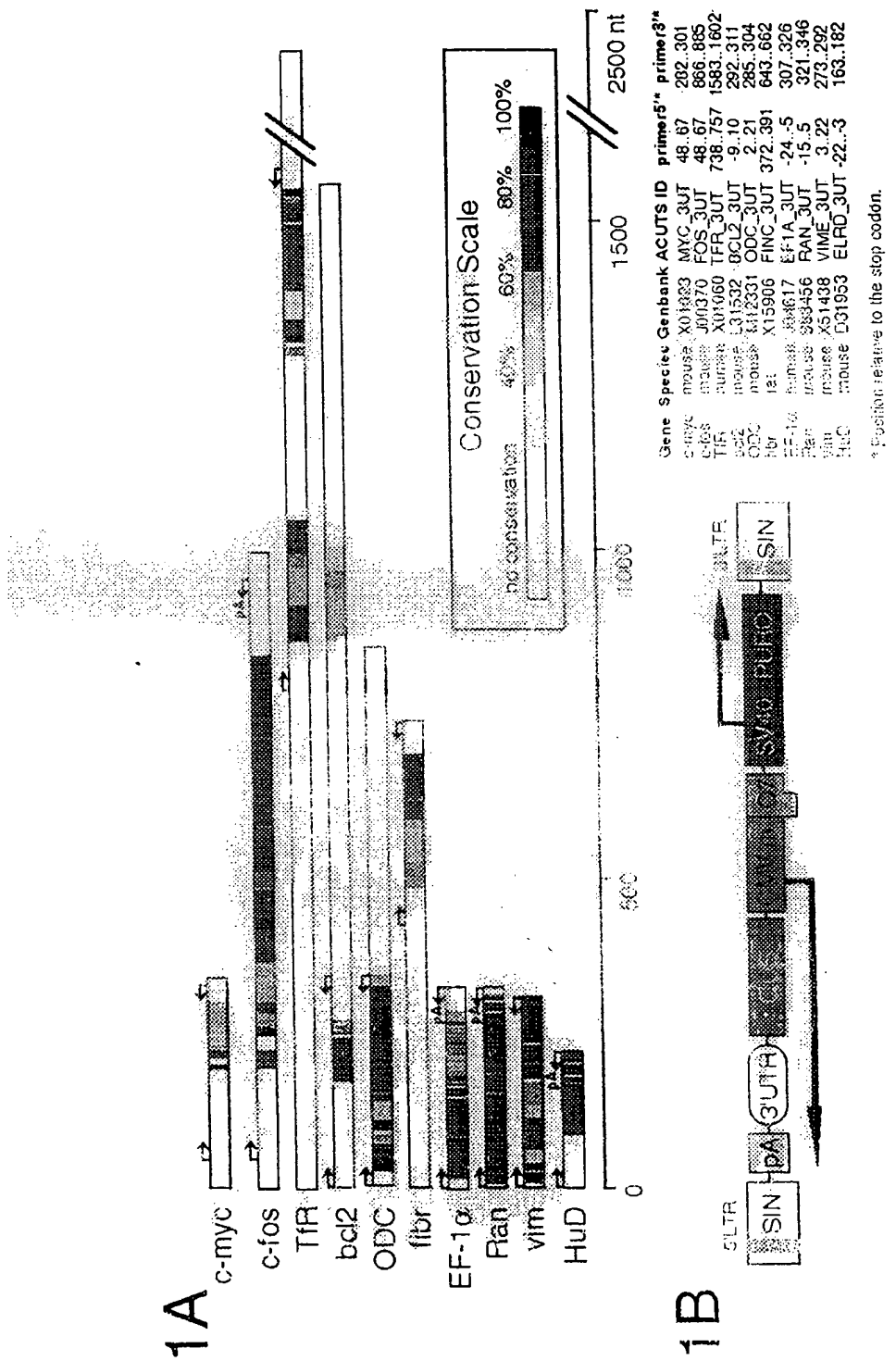
62. The method according to claim 41, wherein the vector is introduced into the host cell *ex vivo*.

63. The method according to claim 41, wherein the vector is introduced into the host cell *in vitro*.

64. The method according to claim 61, wherein the host cell is a mammalian cell.

65. The method according to claim 62, wherein the host cell is a mammalian cell.

66. The method according to claim 63, wherein the host cell is a mammalian cell.

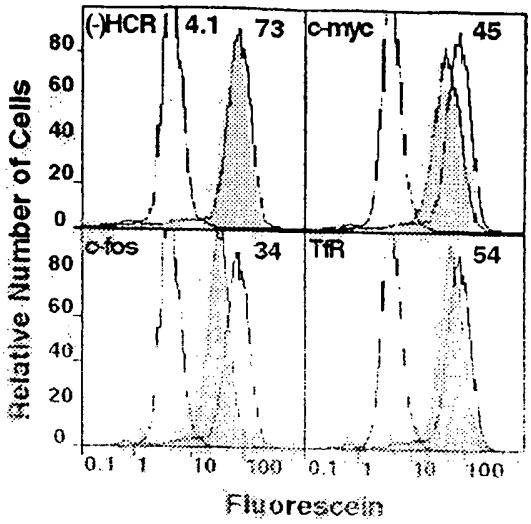


Gene	Species	Genbank	ACUTS ID	primer5'*	primer3'*
c-myc	mouse	X01683	MYC_3UT	48..67	282..301
c-fos	mouse	J00370	FOS_3UT	48..67	866..885
TfR	mouse	X01060	TFR_3UT	738..757	1583..1602
bcl2	mouse	U31532	BCL2_3UT	-9..10	292..311
ODC	mouse	M12331	ODC_3UT	-2..21	285..304
fibr	rat	X15906	FINC_3UT	372..391	643..662
EF-1α	mouse	X04617	EF1A_3UT	-24..-5	307..326
Ran	mouse	393456	RAN_3UT	-15..-5	321..346
vim	mouse	X51438	VIME_3UT	3..22	273..292
HuD	mouse	D31953	ELRD_3UT	-22..-3	163..182

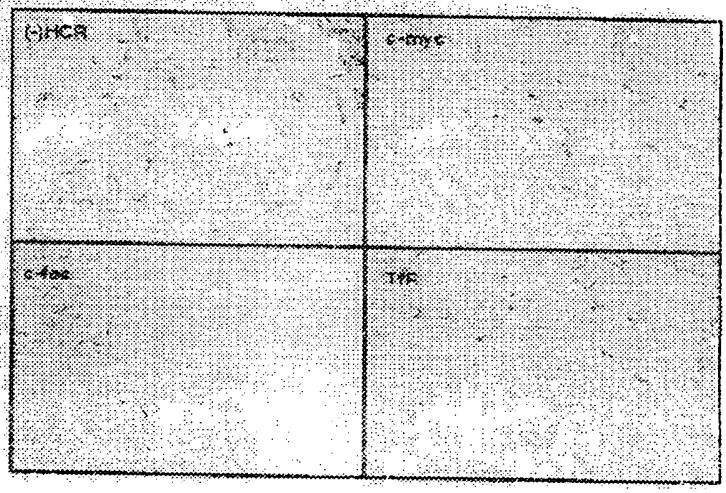
* Position relative to the stop codon.

Figure 1

2A



2B



2C

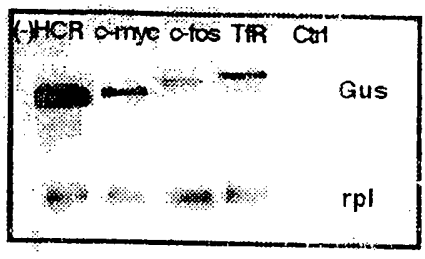


Figure 2

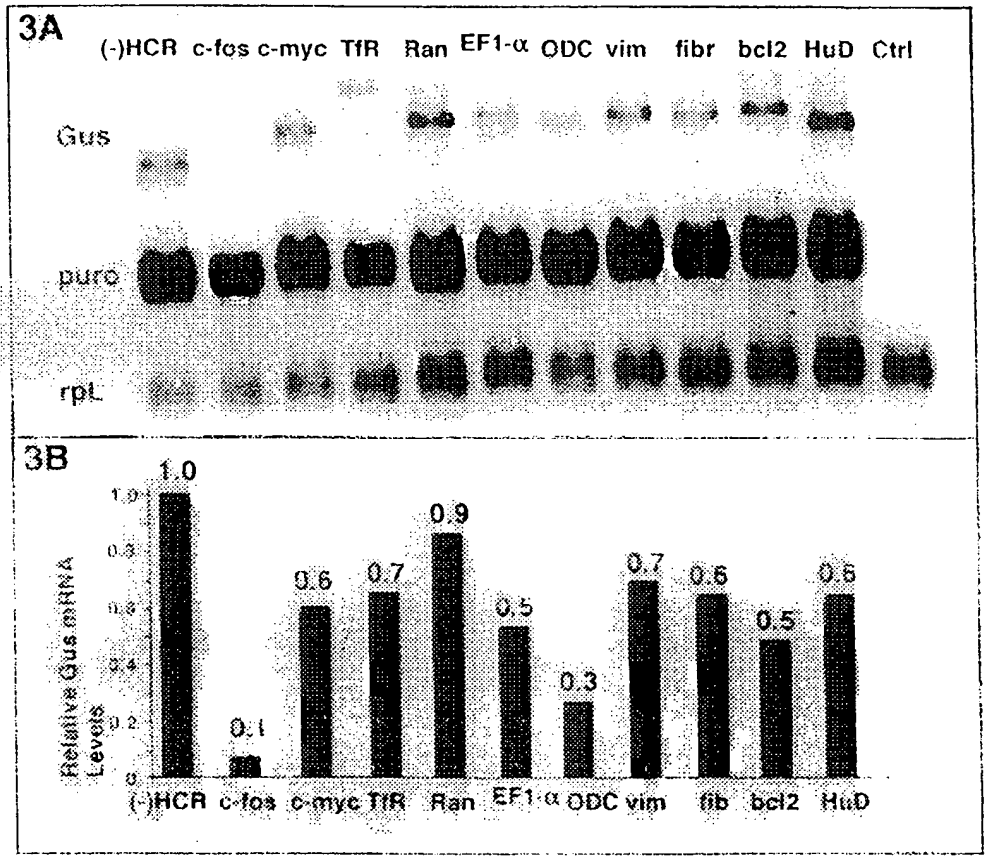

3A
(-)HCR c-fos c-myc TfR Ran EF1-α ODC vim fibr bcl2 HuD Ctrl
Gus
puro
rpL
3B
Relative Gus mRNA Levels
1.0
0.1
0.6
0.7
0.9
0.5
0.3
0.7
0.6
0.5
0.6
(-)HCR c-fos c-myc TfR Ran EF1-α ODC vim fib bcl2 HuD

Figure 3

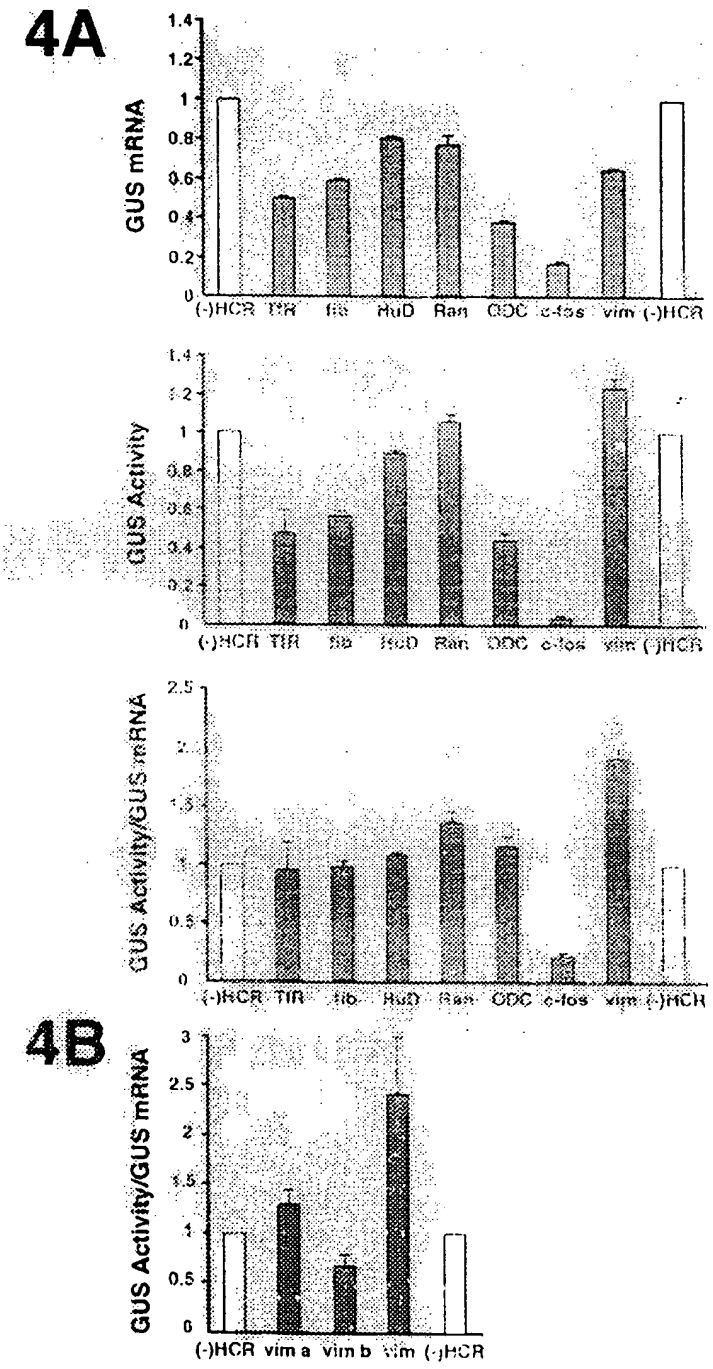

4A
GUS mRNA
1.4
1.2
1
0.8
0.6
0.4
0.2
0
(-)HCR TIR tib HuD Ran ODC c-fos vim (-)HCR
GUS Activity
1.4
1.2
1
0.8
0.6
0.4
0.2
0
(-)HCR TIR tib HuD Ran ODC c-fos vim (-)HCR
GUS Activity/GUS mRNA
2.5
2
1.5
1
0.5
0
(-)HCR TIR tib HuD Ran ODC c-fos vim (-)HCR
4B
GUS Activity/GUS mRNA
3
2.5
2
1.5
1
0.5
0
(-)HCR vim a vim b vim (-)HCR

Figure 4

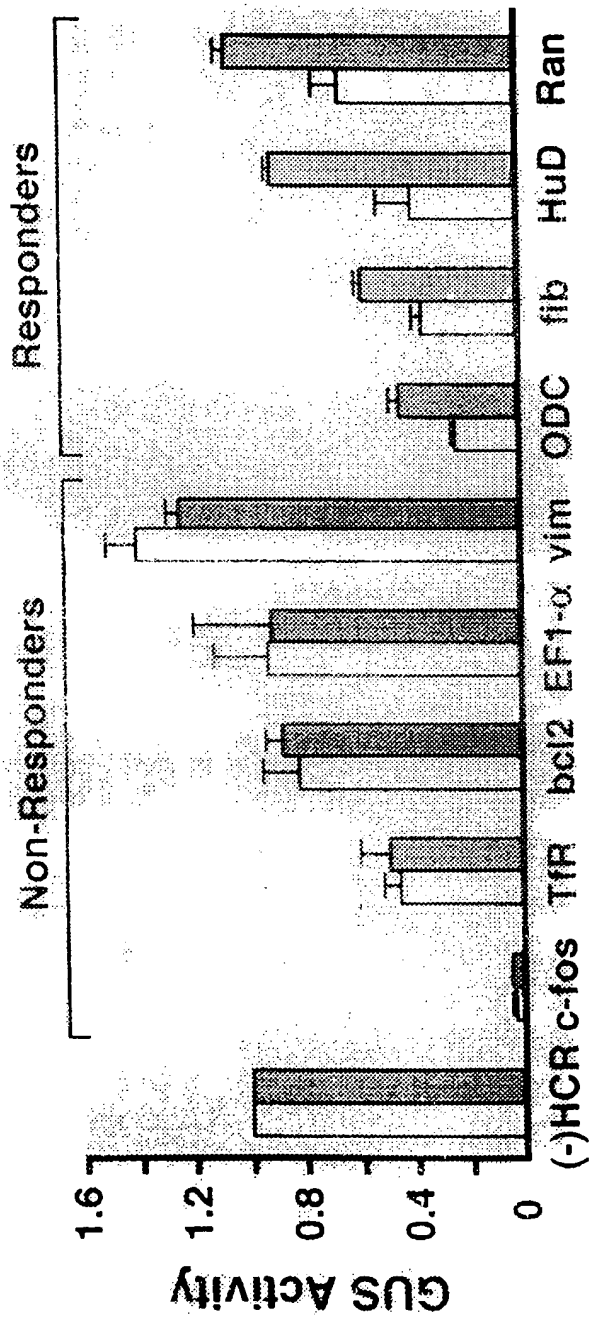

GUS Activity
1.6
1.2
0.8
0.4
0
Non-Responders
Responders
(-)HCR
c-fos
TfR
bcl2
EF1-α
vim
ODC
fib
HuD
Ran

Figure 5

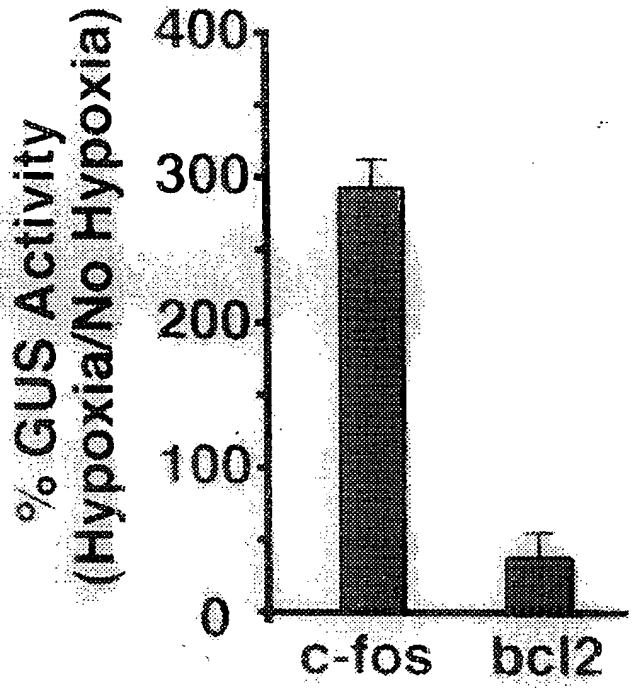


Figure 6

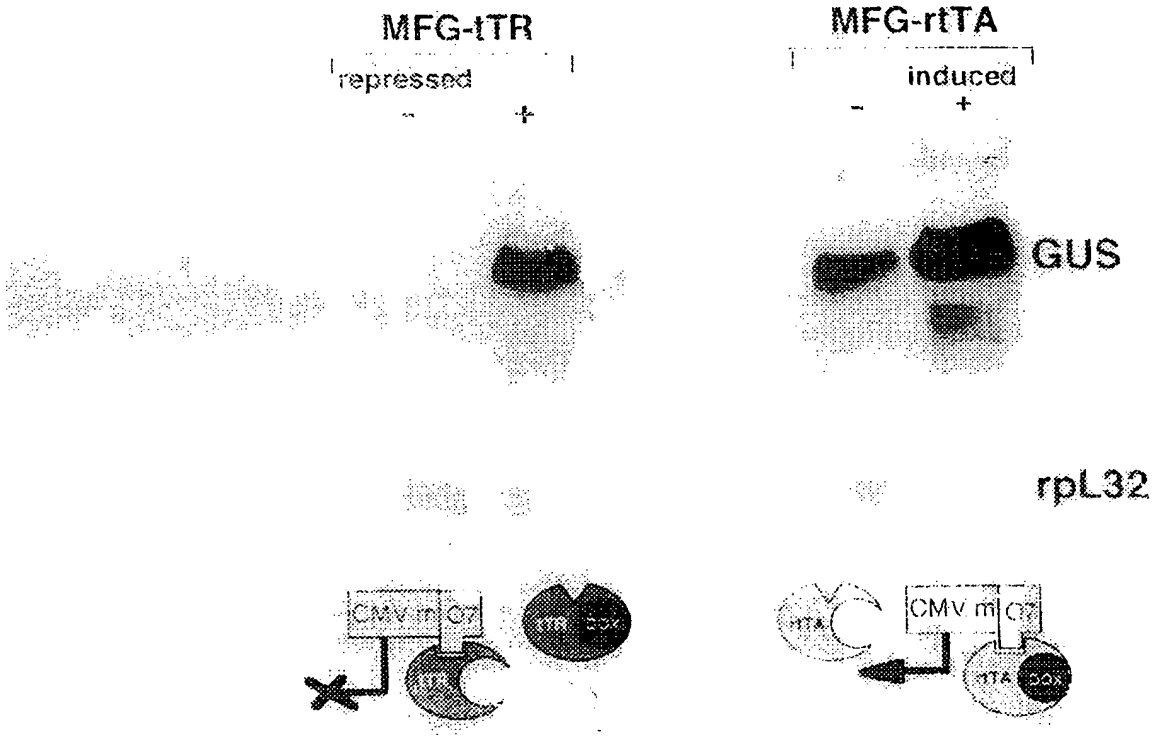

MFG-tTR
repressed
- +
MFG-rtTA
induced
- +
GUS
rpL32
CMV m O2
m
CMV m O2
rtTA
rtTA

Figure 7

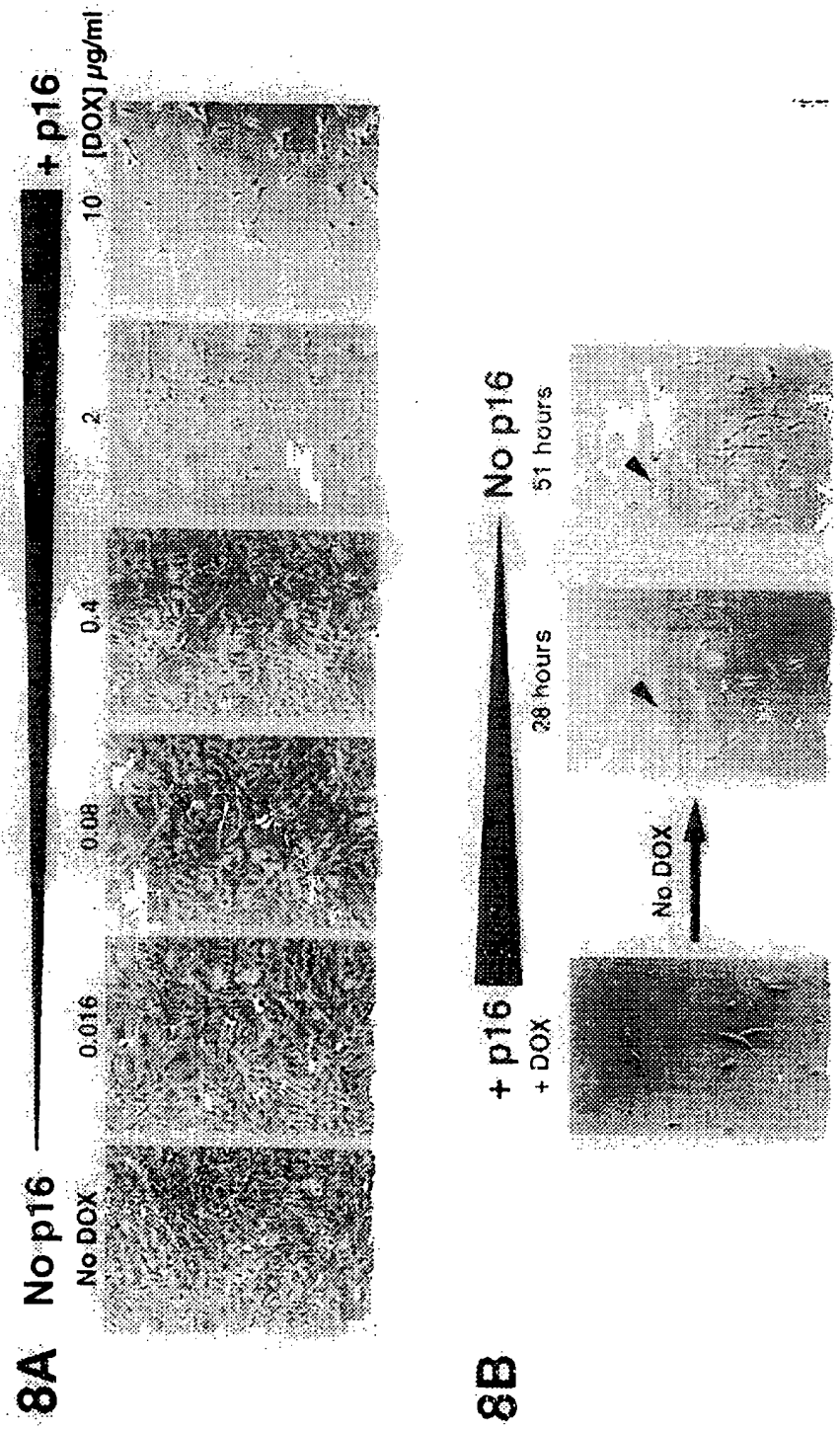

8A
No p16
+ p16
[DOX] μg/ml
10
2
0.4
0.08
0.016
No DOX
8B
No p16
51 hours
28 hours
+ p16
+ DOX
No DOX

Figure 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06093

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 15/85, 5/06; G01N 33/567; C07H 21/00
US CL : 435/320.1, 325, 7.21; 536/24.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1, 325, 7.21; 536/24.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 5,604,118 A (GIRI et al.) 18 February 1997, see entire document.	1 --- 1-38, 40, 52-54, 56, and 60
X --- Y	US 5,589,362 A (BUJARD et al.) 31 December 1996, see entire document.	1, 2, 5, 7, 12, 13, 20, 40, and 52-54 --- 1-38, 40, 52-54, 56, and 60

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
02 JULY 1998	30 JUL 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer CHRISTOPHER S. F. LOW Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06093

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 92/07943 AA (SOMATIX THERAPY CORPORATION) 14 May 1992, see entire document.	1-38, 40, 52-54, 56, and 60
X,P --- Y,P	US 5,654,168 A (BUJARD et al.) 05 August 1997, see entire document.	1, 2, 5, 7, 12, 13, 20, 40, and 52-54 --- 1-38, 40, 52-54, 56, and 60
Y	MORGENSTERN et al. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nuc. Acids Res. 1990, Vol. 18, No. 12, pages 3587-3596, see entire document.	1-38, 40, 52-54, 56, and 60
Y	PFARR et al. Differential Effects of Polyadenylation Regions on Gene Expression in Mammalian Cells. DNA. 1986, Vol. 5, No. 2, pages 115-122, see entire document.	1-38, 40, 52-54, 56, and 60
Y	DURET et al. HOVERGEN: a database of homologous vertebrate genes. Nuc. Acids Res. 1994, Vol. 22, No. 12, pages 2360-2365, see entire document.	1-38, 40, 52-54,56, and 60
Y	US 5,188,828 A (GOLDBERG et al.) 23 February 1993, see entire document.	20, 22, 28, 30, and 38

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06093

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-38, 40, 52-54, 56, and 60

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06093

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

Automated Patent System - USPAT, EPOABS, JPOABS
DIALOG - Files -
Search terms: virus, vector, promoter, long terminal repeat, gene, polyadenylation, regulate, antibiotic, resistance, growth factor(s), proliferation, assay, test, RNA, mammal

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1.

Group I, claims 1-38, 40, 52-54, 56, and 60, drawn to a viral vector (claims 1-37, 40, 53, 56,), cells (claims 52, 54) containing the vectors and a first method (claim 38) of use where candidate RNA is evaluated for functional regulation of transcription.

Group II, claim 39, drawn to a second method where a cDNA library is constructed in a viral vector. The library is not a result of the process of Group I nor is the library necessary to the practice of Group I.

Group III, claims 41-51, 55, 57, and 61-66, drawn to a third method in which expression of RNA is regulated by the presence of a regulatory RNA which is known to regulate transcription both *in vivo* and/or *ex vivo*.

Group IV, claims 58-59, drawn to a fourth method wherein molecules are identified that interact with a known regulatory RNA and the molecules identified thereby,

The inventions listed as Groups I through IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons.

The four groups of claims have produce different results as indicated in the stated groupings. Moreover, the Giri *et al.* (US 5,604,118) reference disclosed viral based vectors where transcription occurred in opposite orientations from different viral regulatory elements and contained a reporter gene. In view of the foregoing, Rule 13.1 is not fulfilled because the special technical features of the vector do not each make a contribution over the prior art.

See Notes to the request form

SUPPLEMENTAL BOX

Continuation of Box No. VI: PRIORITY CLAIM

Item (2): Non-Provisional patent application filed on 26 March 1998 (26.03.98)

TITLE: FUNCTIONAL GENOMIC SCREEN FOR RNA REGULATORY SEQUENCES AND INTERACTING MOLECULES

Attorney Docket No.: 28600-20209.00

Non-Provisional application of: 60/042,543

Express Mail Label No. EM 216 126 962 US